


Research Article

An Intelligent Energy Management Information System with Machine Learning Algorithms in Oil and Gas Industry

Jun Li ¹, Yidong Guo,¹ Zhanbao Fu,² Xiangyang Zhang,² and Fei Shen²

¹Northwest Branch of Research Institute of Petroleum Exploration and Development of CNPC, Beijing 100083, China

²Northwest Branch of Research Institute of Petroleum Exploration and Development of CNPC, Lanzhou 730020, China

Correspondence should be addressed to Jun Li; lijun_xb@petrochina.com.cn

Received 11 August 2022; Accepted 27 September 2022; Published 19 April 2023

Academic Editor: Chia-Huei Wu

Copyright © 2023 Jun Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Oil and gas will continue to play an increasingly important role in global economy development and prosperity for the upcoming years. However, a fact which sometimes is ignored is that the oil and gas industry is an intensive energy-consumption industry and a major contributor to greenhouse gas emissions as well. Recently, with the arrival of the “double carbon” era, the energy-intensive production of oil and gas has caused people to pay more attention. Currently, big data technologies as well as the Internet of things (IoT) are developing at a rapid pace, and the metering infrastructure technology also shows huge advancement; oil and gas companies have built many information systems and developed many functions to acquire the energy consumption and oil production data for the aim of reducing the oil production cost via the emerging information technologies, and it turned out that these investments do improve the energy efficiency and lower oil production cost. Unfortunately, due to the different development standards, some systems store the same data with different values; so, therefore, there is a “data barrier” among these systems and somehow discount the share and analysis of these data resources. To this end, in our work, based on these rich data resources from the other information systems, we discussed an intelligent energy management information system with four machine learning algorithms to enhance the analysis of data resources over energy consumption and oil production. First, we presented the innovation of energy consumption data fusion with the method “Data Lake”; then, four different machine learning algorithms: Support Vector Machine (SVM), Linear Regression (LR), Extreme Learning Machine (ELM), and Artificial Neural Networks (ANN) are installed in the proposed system for predicting the oil and gas production and total energy consumption. In order to aggregate these data into dashboard views that help managers make decisions about operations, we also demonstrated the process of energy consumption data visualization with emerging open-source software tool ECharts. Finally, a real life application of the proposed system was summarized in case of total energy consumption prediction and the system output shows good prediction accuracy which proves the feasibility and benefits of the intelligent energy management information system.

1. Introduction

Oil and gas are generated from decaying plant and animal remains over millions of years. These two types of fuels have played a role as main energy sources for decades. Currently, they are still the primary energy sources for residential lighting, cooking, and heating to transportation and industrial manufacturing. This phenomenon will remain for a long time; in [1], it concludes that the total energy consumption trend will still continue to grow before 2040. However, a fact which sometimes is ignored is that oil and gas companies are

intensive energy-consumption companies and are major contributors to greenhouse gas emissions as well [2]. Recently, with the arrival of the “double carbon” era, the energy-intensive production of oil and gas has drawn a lot of criticism from environmentalist, company managers, investors, and stakeholders as well. Thus, the petroleum industry must pay more efforts on reducing carbon emissions.

Currently, the big data technology as well as the Internet of things (IoT) is developing at a rapid pace, and the metering infrastructure technology also shows a huge

advancement; oil and gas enterprises have invested a lot of efforts on digital software development and hardware construction for the aim of reducing the oil production cost via the emerging information technologies; outcomes show that these investments are beneficial for improving the energy efficiency and lowering the production cost. For example, the WSN has become a widely-used technology in oil and gas fields among the upstream, midstream, and downstream sectors for monitoring the important production processes and controlling the vital operations [3–5]. In [6], it explicitly discussed the advantages of digital technologies of gas production from a different viewpoint, and these promising advantages are included in the case of some special gas wells; with these advanced digital gas exploitation tools, the terms of cost-effective exploitation are most likely to be extended from decreasing production stage, and finally it would contribute to increased labor productivity.

Though, these information technologies bring a lot of benefits for the oil and gas production, and large quantity of data are collected by these devices as well as from thousands of other sensors and systems to help managers make decisions about the operations, often in near-real time. However, due to the different development standards, like data entry name, acquisition frequency, and so on, there is a data barrier between these systems which discounts the usage and analysis of these rich data resources. Therefore, the main question discussed in this study is “how to build an intelligent energy management information system with machine learning algorithms in oil and gas industry to deal with these challenges”; this paper presented the three main functions, including, energy consumption data fusion, energy consumption data analytics, and energy consumption data visualization, of the energy management system in oil and gas industry and shed light on the details of each function as well. The main contribution of this study includes the following:

- (1) Specifically discuss the main functions and the innovations of an intelligent energy management information system in oil and gas industry
- (2) Introduce the flow chart of data fusion with the method “data lake” and the process of data visualization by an open-source software tool ECharts
- (3) Four different machine learning algorithms are installed on this information systems to help enterprise managers or policy makers to analyze the energy consumption data at a company level
- (4) A real life application of this system is analyzed to show how to use this intelligent system to precisely calculate the total energy consumption with the proposed system

The following part of this work is designed as follows: Section 2 describes the related work from the literature review, and the proposed system and its details are discussed in Section 3. Section 4 elaborates the application of the proposed intelligent system, and the conclusion is drawn at the end.

2. Literature Review

In the literature, some intelligent energy management systems based on machine learning methods have been researched; in this section, the related research efforts in this regard are discussed.

Regarding the intelligent energy consumption prediction, various efforts have been done so far. Zhang et al. [7] had introduced an intelligent solar energy system which is able to accurately predict day-ahead power for small-scale solar photovoltaic electricity generators; two commonly-used methods are discussed with good performance and high accuracy. The Rao designed models over the time series for one day-ahead solar power calculation and the best consequence with 9.28% error [8] are presented. In the building area, Guo et al. explored some popular machine learning techniques to forecast total energy consumption over the indoor heating system [9]. In addition, some efforts are made via directly using the regression method to predict the total energy demand; finally, it turned out that the simple regression method like weighted support vector regression and multilinear regression [10–12] are able to achieve promising performance.

In the oil and gas domain, some impressive research work about the intelligent energy management system with digital technologies have been published in [14]; it is concluded that the digital methods have the ability to find the significant opportunities to reform the production condition and enhance the production performance; then, finally, the energy efficiency of oil exploration is improved. And the other related literatures [15–18] also proved that the digital methods are able to offer good chances for the energy reduction over the oil exploration by cutting production cost.

In [13], the author discussed the energy consumption prediction with the commonly-used machine learning algorithms in the oil and gas domain; the results of the experiments show that these intelligent algorithms are able to calculate high prediction result for energy consumption, but the hybrid model of those algorithms would present the better performance; the flow chart of this proposed method is shown in Figure 1. And the proposed method is applied in the energy management system of one oil petroleum company to calculate the total energy consumption amount, and it turns out it has improved the efficiency.

In summary, all these related studies have demonstrated that the intelligent energy management systems have been widely studied and used in industry and achieved good results, and in the oil and gas domain, with the rise of digital technologies, a lot of work about energy prediction has been conducted and presented promising results. Thus, in this paper, we invested the intelligent energy management system with three main functions and four commonly-used algorithms, and a real case of the proposed system was summarized in case of total energy consumption prediction.

3. Proposed System

3.1. The System Architecture. The discussed intelligent energy management system configuration consists of three models: (1) data lake, (2) data analyzing, and (3) data

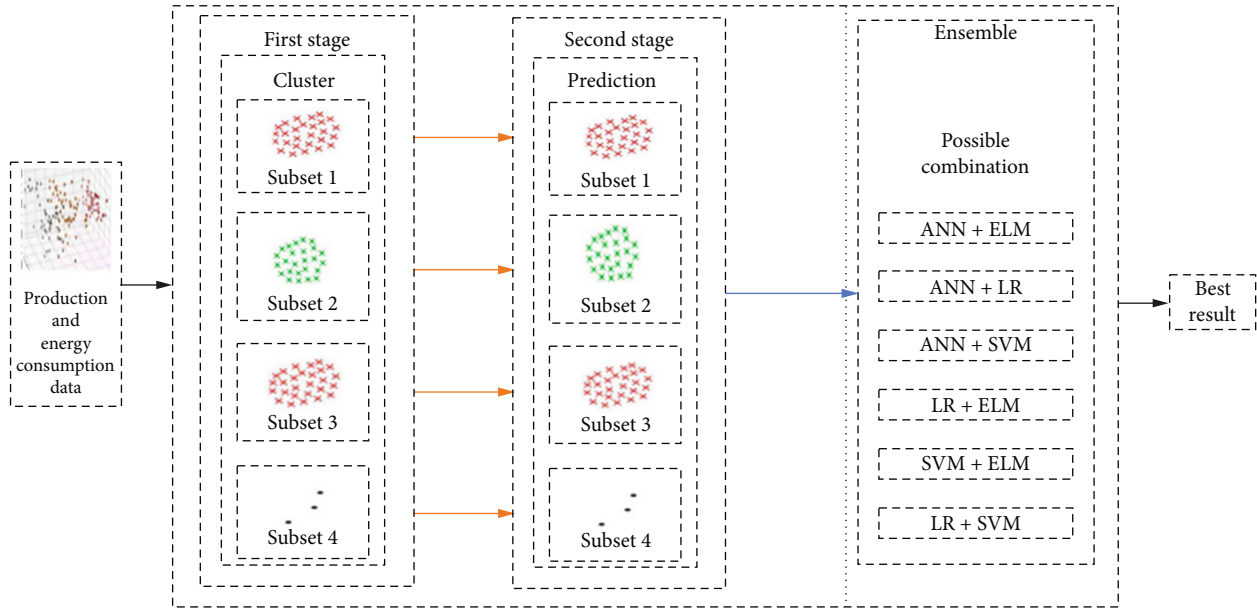


FIGURE 1: The flow chart of the proposed model in [13].

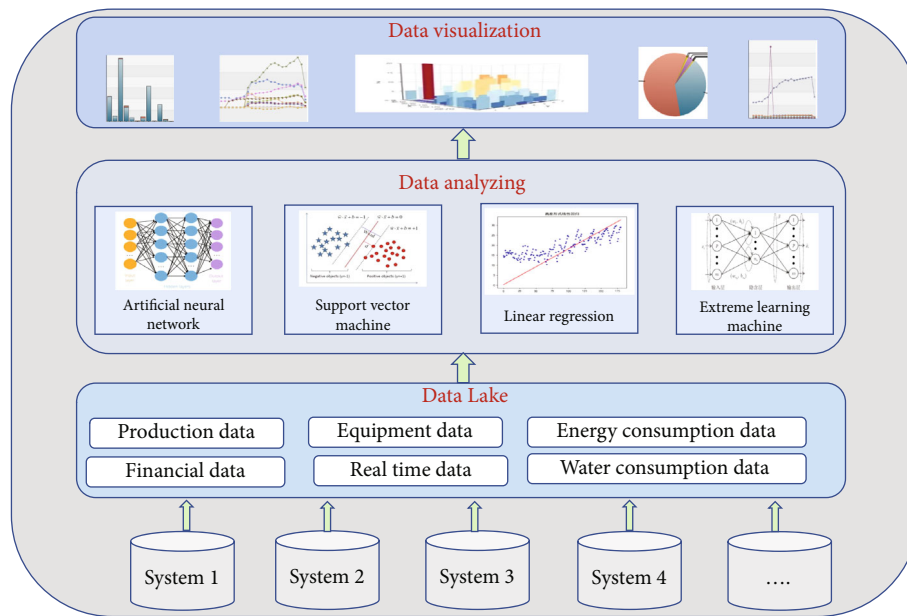


FIGURE 2: The architecture of the proposed system.

visualization. The data lake module presents the data pipe from independent operating systems to the data lake, and then all the data are stored and classified at the data lake level. The data analyzing module describes the data analytics engine with intelligence; four machine learning methods are installed in this engine, and the data from the data lake can use these algorithms directly. The data visualization module shows the data directly with ECharts; in this dashboard the patterns, the trends, and the anomalies are easily identified. The overall architecture is shown in Figure 2.

3.2. Energy Consumption Data Lake in Oil and Gas Field. Currently, the terms of data lake are described as a tool to

accumulate meta data from the other heterogenous data sources. Basically, data lake plays a role of data respiratory in a company, and all kinds of data with different types, format, and structures are stored in it, regardless of structured, semi-structured, or unstructured data [19]. As compared with data warehouse, data lakes are more suitable to store and retrieve various formats of data from the data repository than to play the role of domain specific [20]. Thus, data lakes were built to store all the energy consumption data from the heterogenous data sources, for example, the various system and the related documents of meeting records and training contents.

Guided by the details presented in [20], we proposed an energy consumption data lake system to collect data from

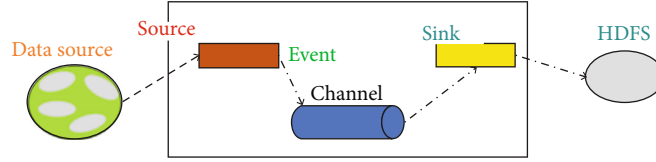


FIGURE 3: The work flow of Apache Flume.

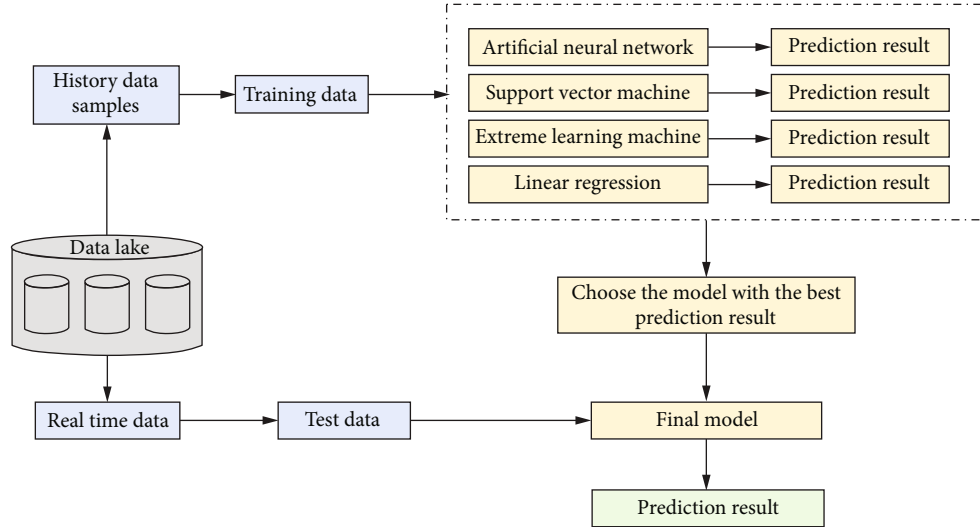


FIGURE 4: The flow chart of data analytics with machine learning.

heterogenous data sources of all related systems in oil and gas fields. The proposed data lake architecture is implemented with Hadoop, which offer the ability to replicate data without high performance calculation and process the data with MapReduce. The proposed data lake system contains four virtual machines which consist of two virtual cores, 32 GB RAM, 500GB disk space.

When the data lake was built, some data technologies to collect data from diverse data source are needed. In the terms of the structured data that are stored in the other related information systems, the Apache Flume tool is mainly considered. This tool has three main components, including, source, sink, and channel, and the work flow is briefly drawn in Figure 3. The source is responsible for collecting the data from the other systems, and then delivers them to the sink for ingestion to the HDFS by the channel. For the other format data, like the data embedded in the HTML pages or PDF documents, some customized scripts are required, and some related work over this issue has been made, like Beautiful Soup 4 (BS4) [21], Selenium [22], and Pandas [23].

We use these tools together to extract the related information from the webpages and the PDF documents of meeting records and training contents.

On the ground of data availability, the manual and automated methods are discussed in our process. For the data which is not often used or updated rarely, like annual reports, we ingested these data manually from the data source into the systems with HDFS commands directly. For the data which is collected every hour, daily or monthly, some scripts are designed to load these data automatically, and in case of the real-time data, a push-based technology

with a custom source in the Flume agent configuration is used to inject the data to the HDFS and to secure reliability; the file channel also which is designed for cached data is considered to handle instances of failure.

3.3. Energy Consumption Data Analytics with Machine Learning. We also develop data processing and analyzing function for the energy management system, and there are various tools available for that, such as Apache Solr and Apache Spark. Apache Spark is popular over data analysis since it offers many APIs to developers for general programming languages, like, Java and Python, and has been widely used in many cases [24]. Moreover, they developed the Spark's machine learning library which includes the generally-used machine learning algorithms and allow the system users to analyze these data in a fast manner. In addition, they also provide the stream processing functions for real-time processing.

Based on the data lake, four commonly-used machine learning methods are installed on the proposed system, so called, support vector machine (SVM), linear regression (LR), extreme learning machine (ELM), and artificial neural network (ANN), and the related work from Section 2 has discuss their ability for prediction. They are developed with the Spark's machine learning library, so that these algorithms are easily configured and designed in our system.

The current implementation of data analyzing with these methods are designed for two cases in our system: (1) predicting the oil and gas production and (2) predicting the total energy consumption. The flow chart for data analyzing with these methods is shown in Figure 4.

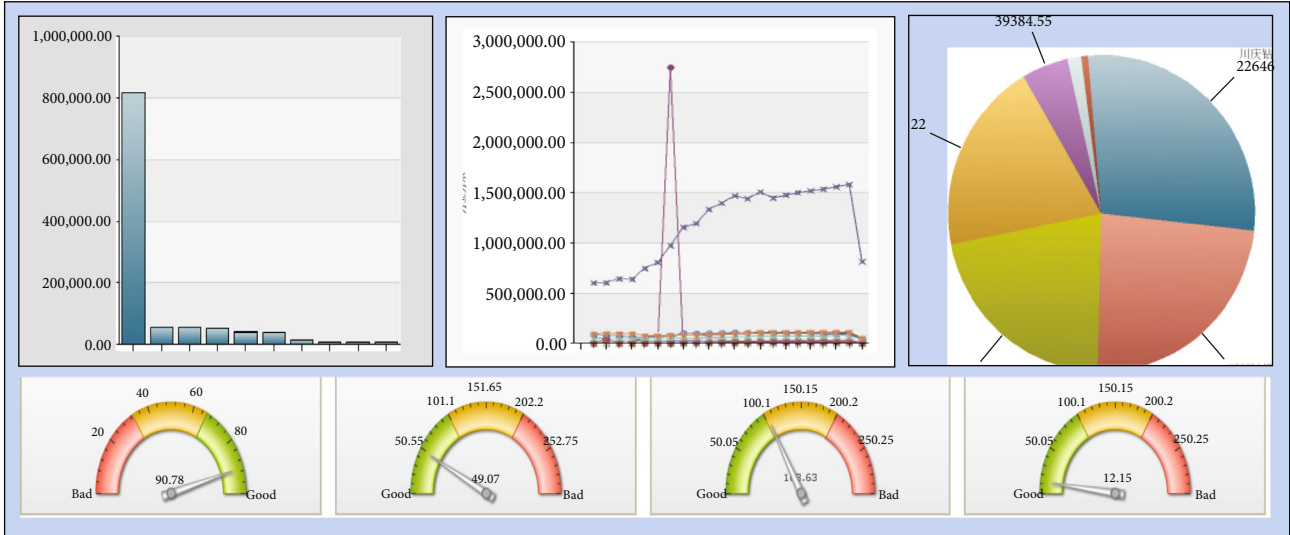


FIGURE 5: The samples of data visualization by ECharts.

For the case of predicting the oil and gas production, first we choose the data from the data lake; the data include the history oil and gas production data; then, we put these history data into these algorithms as training data; when the training procedure is finished, the model with the highest accuracy will be chosen and saved as the final model.

Finally, the final model would be used in practice to help the system users to predict the following days' oil and gas production.

For the case of predicting the total energy consumption, the crude oil output and natural gas output as the feature vector to predict total process-production energy consumption, so the training dataset is represented by tripe $(x, y, \text{ and } z)$, where x means the monthly crude oil output, y denotes monthly natural gas output, and z denotes the monthly total process-production energy consumption. Then, the training data is put into these algorithms for training. When the training process is finished, we mark and save the final model for usage in practice.

3.4. Energy Consumption Data Visualization by ECharts. The data analyzing has presented some insights of data from data lake, but for more insight and details from the data, the data visualization turns out to be a powerful tool for that, and it can easily be developed through web-interface or the Java-based and python-based library, which allows developers to design bar graph, pie charts, and the other dynamic dashboards with graphics very quickly.

In our study, the most widely-used tool ECharts [25] is discussed. The ECharts does supply almost all the basic graphs and charts that the users needed for data visualization, like, line charts, bar charts, scatter charts, and so on. What is more interesting is that developers are able to develop charts and graphs with ECharts very easily and quickly without too much software developing knowledge. In addition, ECharts can display the visualization of tons of data with extremely large volume, and it would not discount the ability of scaling and transforming of data. Con-

cerning the friendly interface and the geographic data, ECharts offers eye-catching visual effects for the line data and point data visualization. And it turns out to be a very declarative framework for rapid construction of web-based visualization. The samples of data visualization are shown in Figure 5.

4. Use Case of our System

In this part, we will use the dataset collected from one oil and gas company's energy management system to validate the performance of our proposed system. The flow process consists of three main steps, including, dataset, model design, and prediction result.

4.1. Dataset. In our work, we collected the data from the company's energy management system; according to the requirements of the system management, the energy related data are collected once every month. Thus, we considered a ten-year dataset which consist of 120 observations in total. Part of the data samples is shown in Figure 6, and there is strong complexity and weak time relation between these data samples. Thus, the time-related parameters are not considered in our paper and the tripe $(x, y, \text{ and } z)$ is designed; where x means the monthly crude oil output, y denotes monthly natural gas output, and z denotes the monthly total process-production energy consumption. These 120 data samples are divided into training dataset and test dataset, and the training dataset have 90 data samples and the rest of data belongs to the test dataset.

4.2. Model Design. In our intelligent energy management systems, there are four machine learning algorithms installed which included SVM, LR, ELM, and ANN.

For the LR model, concerning the dimension of the input, there are simple linear regression and multiple linear regression; in our work, the input vector has two dimensions, namely, crude oil output and natural gas output, so

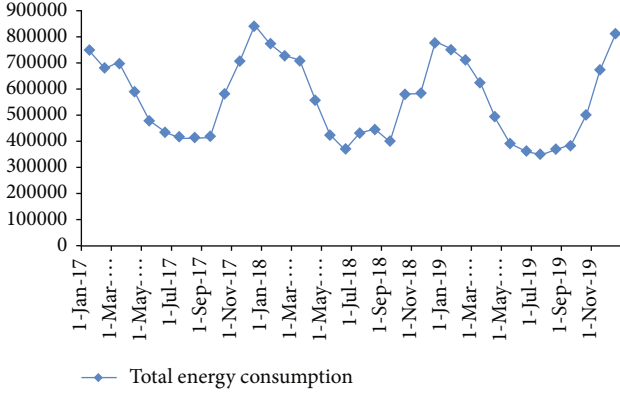


FIGURE 6: Part of energy production and consumption data sample.

the multiple linear regression model with equation (1) is discussed:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2. \quad (1)$$

Where Y denotes the prediction result, X_1 and X_2 mean the monthly crude oil output and monthly gas production, and β_0 , β_1 , and β_2 are the coefficient of the multiple linear regression model. For the ANN, due to the limitation of the dataset size, a three-layer artificial neural network is built in the system to forecast the monthly energy consumption, and the first layer has 2 neurons, the hidden layer has 3, respectively, and the final layer has of 1 neuron for the total energy consumption prediction result. The sigmoid is chosen as the active function between these layers. Compared with ANN, the ELM just contains one hidden layer, so, therefore, highly reduces the computational cost. Regarding the energy prediction, a lot of research work has been done with this model; in this paper, we build the ELM model which consists of 4 neurons in the middle layer to forecast the energy consumption. Finally, the SVM also is considered to validate the predicting performance of the monthly energy consumption prediction, and we choose polynomial kernel function from the model to construct the SVM model. All the design and training procedure are finished through the intelligent energy management system.

4.3. Prediction Result. This part would discuss the prediction result of the installed four machine learning methods. We chose the RMSE as the metric to evaluate the prediction performance of these models:

$$\text{RMSE} = \sqrt{\frac{\sum_i^N (v_i - v'_i)^2}{N}}, \quad (2)$$

where N means the size of data set, v_i denotes the label value, and v'_i means the calculated value. Usually, the better method prediction performance is presented with smaller value of RMSE. Table 1 shows the RMSE result of four methods over training and test dataset.

TABLE 1: The RMSE result of four models over training and test dataset.

Model	Test dataset	Training dataset
Support vector machine	4.23	3.32
Linear regression	6.73	2.24
Extreme learning machine	0.78	0.37
Artificial neural network	0.63	0.42

These experiment results show that the four methods are able to calculate the prediction results with good accuracy. However, the LR method achieved prediction results with the lowest accuracy, and it revealed that there is no strong linear correlation between the oil production and energy consumption. The ELM algorithm and the ANN predicted the energy consumption with highest accuracy, and it revealed that these two models can be used to forecast monthly total energy consumption in a company and demonstrated the feasibility of the management system for the case of monthly total energy consumption calculation in oil and gas sector.

5. Conclusion

Many oil and gas companies have collected a lot of data and these data are lacking future analyzation. Thus, in this paper, we discussed the intelligent energy management system in the oil and gas industry. First, the innovation over breaking the “data barrier” among each system was introduced; we discussed the methodology of “data lake”. Then, aimed at enhancing data analytics and making most of these data, four different machine learning algorithms, SVM, LR, SVM, and ELM, are installed in the proposed system with the “data lake” for predicting the oil and gas production and total energy consumption; the system offers the whole process of model design, training, and calculation. The system can make use of such models directly without any code work. In order to visualize these data and show the data trends on the dashboard, an emerging open-source software tool ECharts was presented. Finally, we applied the proposed system into one oil and gas company to validate the feasibility of this system, and it suggests that this system is able to calculate the prediction results with good accuracy, but due to no strong linear correlation between these data, the artificial neural network achieves the best result.

In future, the primary task is to improve the system’s performance security, scalability, accuracy, standardization, automation, and some other aspects. To achieve this goal, we will try to use more data from other oil and gas companies to validate the systems’ performance and robustness, and more intelligent methods would be considered to be installed in this system. What is more, combining the proposed system with the other big data and advanced new IT technologies to develop more intelligent energy management information system is our main upcoming work as well.

Data Availability

All dataset used during the study has appeared in the submitted article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] 2019, <https://www.bp.com/en/global/corporate/energyeconomics/energy-%20outlook/introduction/overview%20w.html>.
- [2] H. Saadawi, "Application of renewable energy in the oil and gas industry," in *SPE Middle East Oil and Gas Show and Conference*, Manama, Bahrain, 2019.
- [3] S. Savazzi, U. Spagnolini, L. Goratti, D. Molteni, M. Latva-aho, and M. Nicoli, "Ultra-wide band sensor networks in oil and gas explorations," *IEEE Communications Magazine*, vol. 51, no. 4, pp. 150–160, 2013.
- [4] C. M. Corneliu, D. Valentin, and S. Valentin, "Optimal location for underwater data collection sensor network nodes creating multi-purpose "safe-nets" around offshore exploration areas," in *2013 17th International Conference on System Theory, Control and Computing (ICSTCC)*, pp. 601–606, Sinaia, Sinaia, Romania, 2013.
- [5] S. Savazzi and U. Spagnolin, "Synchronous ultra-wide band wireless sensors networks for oil and gas exploration," in *2009 IEEE Symposium on Computers and Communications*, pp. 907–912, Sousse, Tunisia, 2009.
- [6] A. N. Dmitrievskiy, N. A. Eremin, and V. E. Stolyarov, "Digital transformation of gas production," *IOP Conference Series: Materials Science and Engineering*, vol. 700, no. 1, article 012052, 2019.
- [7] Y. Zhang, M. Beaudin, R. Taheri, H. Zareipour, and D. Wood, "Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2253–2262, 2015.
- [8] V. Prema and K. U. Rao, "Development of statistical time series models for solar power prediction," *Renewable Energy*, vol. 83, pp. 100–109, 2015.
- [9] Y. Guo, J. Wang, H. Chen et al., "Machine learning-based thermal response time ahead energy demand prediction for building heating systems," *Applied Energy*, vol. 221, pp. 16–27, 2018.
- [10] F. Zhang, C. Deb, S. E. Lee, J. Yang, and K. W. Shah, "Time series forecasting for building energy consumption using weighted support vector regression with differential evolution optimization technique," *Energy and Buildings*, vol. 126, pp. 94–103, 2016.
- [11] S. Asadi, S. S. Amiri, and M. Mottahedi, "On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design," *Energy and Buildings*, vol. 85, pp. 246–255, 2014.
- [12] N. Fumo and M. A. Rafe Biswas, "Regression analysis for prediction of residential energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 47, pp. 332–343, 2015.
- [13] J. Li, Y. Guo, X. Zhang, and Z. Fu, "Using hybrid machine learning methods to predict and improve the energy consumption efficiency in oil and gas fields," *Mobile Information Systems*, vol. 2021, Article ID 5729630, 7 pages, 2021.
- [14] C. Temizel, C. H. Canbaz, Y. Palabiyik et al., "A comprehensive review of smart/intelligent oilfield technologies and applications in the oil and gas industry," in *SPE Middle East Oil and Gas Show and Conference*, Manama, Bahrain, 2019.
- [15] M. R. Brulé, *Group IBMS the Data Reservoir: How Big Data Technologies Advance Data Management and Analytics in E & P Introduction—General Data Reservoir Concepts Data Reservoir for E & P*, 2015.
- [16] K. Hilgefert, "Big data analysis using Bayesian network modeling: a case study with WG-ICDA of a gas storage field," in *CORROSION 2018*, Phoenix, Arizona, USA, April 2018.
- [17] J. Sun, Z. Zhang, and X. Sun, "The intelligent crude oil anti-theft system based on IoT under different scenarios," *Procedia Computer Science*, vol. 96, pp. 1581–1588, 2016.
- [18] W. Z. Khan, M. Y. Aalsalem, M. K. Khan, M. S. Hossain, and M. Atiquzzaman, "A reliable Internet of things based architecture for oil and gas industry," in *2017 19th International Conference on Advanced Communication Technology (ICACT)*, pp. 705–710, PyeongChang, Korea (South), 2017.
- [19] P. P. Khine and Z. S. Wang, "Data lake: a new ideology in big data era," *EDP Sciences*, vol. 17, article 03025, 2018.
- [20] H. Mehmood, E. Gilman, M. Cortes et al., "Implementing big data lake for heterogeneous data sources," in *2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW)*, pp. 37–44, Macao, China, 2019.
- [21] *beautifulsoup4*PyPIJanuary 2019, <https://pypi.org/project/beautifulsoup4/>.
- [22] *Selenium Blog Posts*January 2019, <https://www.seleniumhq.org/>.
- [23] *Python Data Analysis Library*January 2019, <https://pandas.pydata.org/>.
- [24] S. Salloum, R. Dautov, X. Chen, P. X. Peng, and J. Z. Huang, "Big data analytics on apache spark," *International Journal of Data Science and Analytics*, vol. 1, no. 3-4, pp. 145–164, 2016.
- [25] <https://echarts.apache.org/en/feature.html>.