ISSN: 2096-3246
Volume 54, Issue 02, April, 2022

Big Data Interpretation for Petroleum and Energy Extraction Study and Application

V. Naveen Kumar[1] , Dr. Ashok Kumar.P.S[1]
Bapatla Engineering College, Bapatla, Andhra Pradesh[1]
Don Bosco Institute of Technology, Karnataka[2]

Keywords:
Oil Production, Big Data, Application, Spark.

ABSTRACT

The exploration and commercialization of petroleum and energy is a data-driven economy. The generation of massive volumes of data has laid the ground- work for the application of Big Data Analytics (BDA), mining data resources, and using this information to drive oilfield production techniques are all part of the core technologies of the oil and gas industry. Huabei Oilfield has been experimenting with the use of big data analytics in oil and gas production in recent years. Taking into account the various categories and attributes of oilfield information, a "seven-step method" for closed-loop BDA has been developed, which includes data collection, processing, tracking, and evaluation. A Hadoop/Spark-based data-mining platform for oil production engineering has also been developed.

1. INTRODUCTION

Massive data, the cloud, and artificial intelligence (AI) have emerged as the fundamental propelling factors behind modern industry's transformation and growth, as well as the improvement of product excellence and productivity. Huabei Oilfield has been committed to the development of oil field information in recent years [1]. They built China's initial computerized coal-bed natural gas pipeline. Efforts have also been undertaken to create smart oilfields in the traditional oil and gas production sec- tor. 70% of production data in today's oil wells is gathered automatically; smart identification and analysis; customized surveillance monitoring; distant control of pumps; intelligent gathering and conveyance; and power outage alerting. BDA in the oil and gas sector will be aided by the water system's ability to collect and monitor data automatically [2]. As part of its exploratory study into the use of BDA in major do- mains of oil and gas production, Huabei Oilfield has already obtained early application results in the areas of exploration and development, engineering and management. An oilfield production industry-specific "seven phase" closed-loop BDA framework and processes, as well as an oilfield corporation level information extraction framework, have been developed to provide scientific means for developing specialized strategies for power savers and usage lowering, productivity improvements, and safe operation [3].

2. RELATED WORKS

The "seven step" method of closed-loop BDA from gathering, processing, monitoring, and assessment is designed in accordance with the needs of the petroleum and energy production industry and the peculiarities of data type [4]. Having a clear goal for BDA is critical. With data analysis value and potential (such as model performance and power usage) in mind, it is important to establish a goal based on what is really being produced in order to drive the actual production process. It is also necessary to establish data collection conditions for the aim of the analysis, so that the relevant data needed for analysis may be gathered in accordance with the oilfield information construction standards [5]. Collection, screening, and cleanup are all major components of a data cleansing process that also includes formats like CSV and XML. All important aspects, for instance, can be obtained from conceptual expertise or operational expertise. Then, using the analytical markers as a guide, gather all pertinent data items and define the data type and preservation position for each data item. Perform the preliminary work for data collecting. Select analysis indicators according to the analysis objectives [6]. Engineering data has a number of flaws, including data omissions, duplication, errors, inconsistencies in format, and a lack of uniformity among units. The integrity and accuracy of the findings will be affected to some degree because of this. Before analysing the acquired data, it is necessary to cleanse it. Screening, cleaning, integrating, formatting, and other similar activities are all part of data processing. It's important to standardise and clean up each piece of information in order to verify that the data is accurate and ready for mining [7].

Analysis indexes are mined once the data is cleaned, and each data item and analysis index are mined for their impact laws. It is possible to classify and predict time series data using a variety of different data mining techniques, including regression and cluster analysis. Standard mathematical analytics, conformation analytics, relational analytics, categorization analytics, regression analytics, etc. are some of the most often used approaches for analysing huge data in oil production engineering. When it comes to data mining, statistics and comparative analysis are the most often employed tools [8]. The purpose of categorization analytics is to narrow the space between species and make the law clear in certain cases when the effect of the law is less obvious. The goal of relevance analysis is to discover the most important component in the connection between two or more events. In most cases, regression analysis is utilised to determine the relationships between variables and the resulting indices [9].

The ability to see the outcomes of data mining is critical, as it aids analysts in dis- covering patterns and spotting possible issues. Graphics and tables are required be- cause data mining findings have to be presented visually. Analysis of mining analysis results using a visual chart is used to find and evaluate issues, as well as their causes and the primary elements influencing them [10]. This information is then used to assist design a better mining strategy, based on the findings of diagnosis and analysis. Diagnosis findings are used to create the corrective procedures and to improve process parameters that will have an impact on performance. The adjustment plan is put into action on the ground, the findings of the analysis are in, and the effects are being tracked and evaluated [11]. Collect data before and after the scheme is implemented, evaluate the impact and benefit of the implementation, and then examine the data collected after the scheme is implemented to create a closed-loop study [12].

3. BIG DATA MINING PLATFORM

To promote the utilization of big data mining in petroleum and energy development and to overcome the problems of real-time data and large information, we are building a Big Data Mining Framework (BDMF) for petroleum and energy development. Big data mining is a broad quantitative solution that integrates information computation, pattern development, and pragmatic implementation depending on an information resource pool. All the capabilities you need to perform a variety of statistical analyses and visualizations may be found in this one software package. Model generation and release may be done dynamically. Hierarchical design is used in the architecture of the petroleum and energy production BDMF, which consists of the following layers: integration and cleaning; storage; computation; mining (offline/online)/application; and a data mining layer (offline/online).

The information relation, information cleansing, information extraction, and data visualisation modules are the primary components of this modular big data mining platform. It has a user-friendly UI that is clever. A wide range of frequently used data mining methods and parameters may be applied to the platform, which allows for the storing and administration of many different data sources, as well as a wide range of data cleaning procedures. Dragging, pulling, and joining techniques are supported by the platform for process modification. Security and dependability are ensured by relying on authority authorisation. Model release is supported as well. The hydrostatic performance of pumping wells is used as an illustration of big data mining in this study. Wellbore conditions worsen, oil well liquid output is decreased more and more each year as the oilfield progresses to its medium and late stages of development, and the number of low-efficiency wells grows year after year. Pump efficiency may be improved and energy usage decreased by examining the pump's sensitive variables. For starters, a theoretical study of pump efficiency at a single well is conducted to identify the influencing factor set, and preliminarily cleaned data is collected.

As of right now, tubular pumps are the most often utilized pump in oil field pumping wells. Pump efficiency η refers to the ratio of real liquid production Q to theoretical liquid production Q_t, which is often lower in pumping wells. The efficiency of a single well pump is affected by a total of 11 parameters, including daily fluid output, which are influenced by three characteristics of the equipment, operating system, and formation. Form: real consistent watery generation, water cut, submergence, pump depth, pumping well variance and viscosity; operating attributes: stroke, frequency and load; device parameters: hydraulic circumference and conceptual consistent liquid production are just a few of the variables that must be considered. The BDMF is used to find the most essential control affecting aspects via the generation of a data collection and the analytical technique for grey similarity analytics of piston effectiveness variables. Analytical concepts are shown in Fig.1. A1, A2, and other systems are the primary sources of pump efficiency altering variables. In order to build a dataset, merge statements from the hive. Build an analytical process: remove any missing values; normalise the data; remove the impact of the dimension (unit); then do a grey correlation analysis on the data set. You may see the outcomes of your analysis by submitting the work and seeing the results.

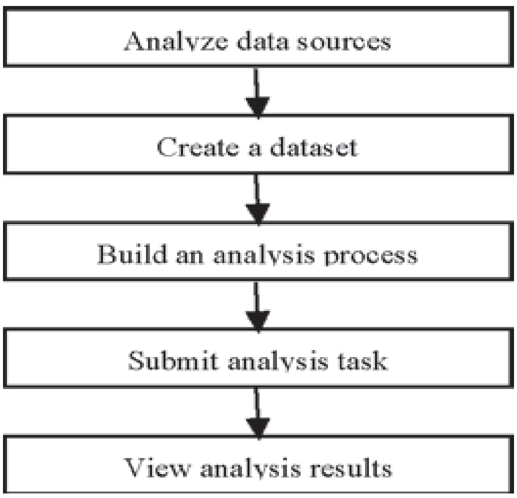

Fig.1 Analysis of the primary controlling variables.

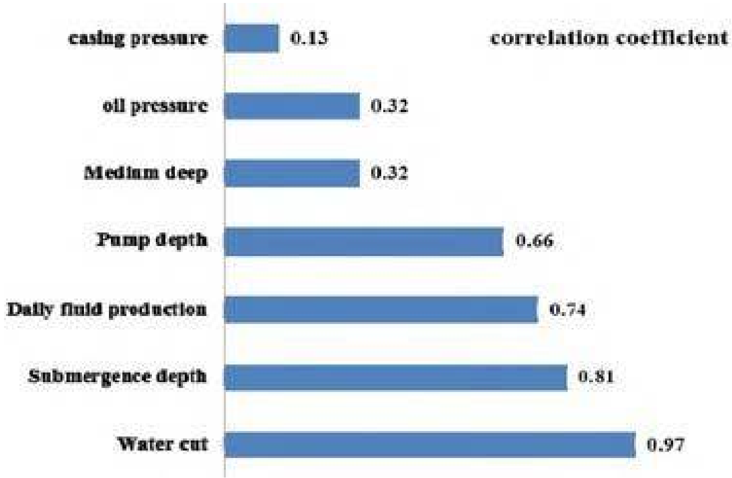

Fig.2 A graph showing the correlation coefficients of the many elements that affect pump performance.

4. RESULTS AND DISCUSSION

The "oil and gas production data mining framework" extensively examines the key deciding aspects of pumping system efficiency, such as water cut, subsidence, every- day hydraulic output, and shaft profundity. The matching framework of pumping system performance and everyday hydraulic generation in Huabei Oilfield is depicted in Fig. 3 utilizing two-dimensional data visualization and polynomial matching mechanism, and using the analysis of pumping system efficiency and everyday water productivity as an illustration. Additionally, a forecast model for daily fluid output and pump efficiency has been developed. R^2=0.736343 is the fitting percentage for them. When the matching level is greater than or equal to 0.3, the matching and similarity are considered acceptable by the regression analysis judgment criteria [13]. Several other major control parameters are fitted and assessed using the same methodology.

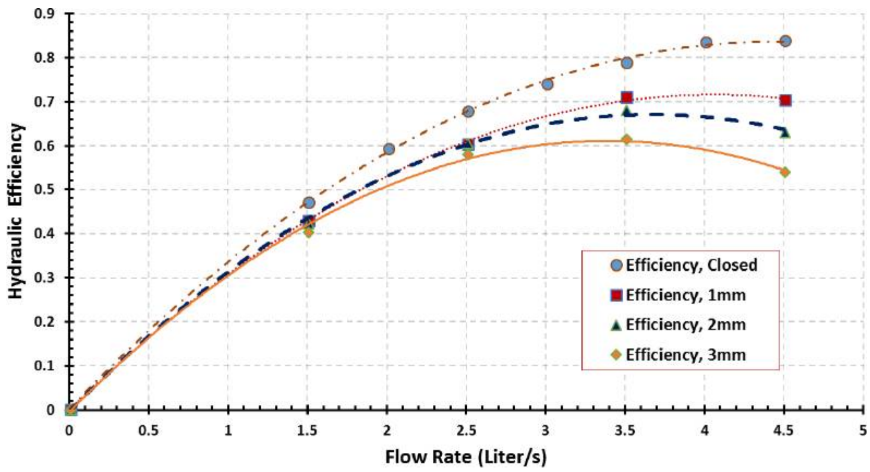

Fig.3 Pumping system performance and everyday liquid production matching diagram.

The univariate matching hypothesis that determines pumping effectiveness is shown in Table 1. The primary influencing elements of pump efficiency are discovered, laying the groundwork for pump efficiency enhancement. Since 2015, Huabei Oilfield has promoted and used BDA technology in several industries, including petroleum production, water injection, gathering, and transportation. Huabei Oilfield views the energy efficiency analysis of pumping wells as a breakthrough. A total of 2560 wells have been improved and modified for pumping wells since 2015, resulting in an in- crease of more than 50,000 tonnes of oil and a 124 million yuan economic gain. Fol- lowing deployment, the pumping well system's efficiency rose by 2.3 percent, and a 100-meter tonne of fluid now uses 0.12 kW less electricity per hour.

Table 1. Huabei Oil Field Pump Efficiency Single Factor Prediction Model

Impact factor	**Forecasting Model**	R^2
Humidity level	$\eta = 20.143 * e^{0.008}$	0.98
Subduction	$\eta = -0.176\ \Delta H^2 + 3.654\ \Delta\ H + 36.932$	0.82
Regular Water	$\eta = \log(Q_t) + 1.7654$	0.75
Profundity of the pump	$\eta = -0.77\ H^2 + 5.86\ H + 14.96$	0.67

5. CONCLUSION AND FUTURE SCOPE

BDA motivates information utilisation and successfully recognises prospective deliberate valuation legislation in production systems, offering new efficacious ways to mine output prospects, directing and optimising manufacturing execution variables, and supplying technological assistance for accomplishing energy savings, utilisation diminution, and efficacious governance. BDA technology exploration and in- novation for the oil and gas sector may provide total operational observation, self- diagnosis of faults, method refinement, automatic variable modification, and other capabilities that can further improve oilfield production administration. Data is the foundation of BDA. For the present oilfield data resource development, new specifications have been put out on how to guarantee that data construction is practical, effective, accurate, comprehensive, safe, and unified.

5. REFERENCES

[1] M. Guizhi, L. Zhongbo, Z. Zhanmin, J. Xuefeng, X. Weiyi and W. Shuai, "Research and Application of BDA for Oil and Gas Production," 2021 IEEE 2nd International Confer- ence on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), 2021, pp. 97-100, doi: 10.1109/ICBAIE52039.2021.9389892.

[2] R. M. Alguliyev, R. M. Aliguliyev and M. S. Hajirahimova, "Big data integration architectural concepts for oil and gas industry," 2016 IEEE 10th International Conference on Ap- plication of Information and Communication Technologies (AICT), 2016, pp. 1-5, doi: 10.1109/ICAICT.2016.7991832.

[3] R. N. Anderson, "'Petroleum Analytics Learning Machine' for optimizing the Internet of Things of today's digital oil field-to-refinery petroleum system," 2017 IEEE International Conference on Big Data (Big Data), 2017, pp. 4542-4545, doi: 10.1109/BigData.2017.8258496.

[4] M. Riasetiawan, F. Anggara, A. Ashari, S. Winardi and B. N. Prastowo, "Data Model and Analysis for Big Data Mapping and Management in the Energy Data Platform," 2021 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA), 2021, pp. 1-6, doi: 10.1109/DATABIA53375.2021.9650117.

[5] A. Eryilmaz, R. Srikant and J. R. Perkins, "Stable scheduling policies for fading wireless channels", IEEE/ACM Trans. on Networking, Vol. 13, 2005, pp.411-424.

[6] D. Juan, W. Yue, J. Wei and L. Shuyuan, "Research on the Early-warning Model of Foaming Phenomenon in Desulfurization Solution System of Natural Gas Purification Plant Based on Artificial Intelligence Technology," 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), 2019, pp. 215-222, doi: 10.1109/ICAIBD.2019.8836996.

[7] L. Jiadi, H. Yang, L. Huan, Z. Xinli and L. WenJing, "Research on Data Center Operation and Maintenance Management Based on Big Data," 2020 International Conference on Computer Engineering and Application (ICCEA), 2020, pp. 124-127, doi: 10.1109/ICCEA50009.2020.00033.

[8] L. Li and Y. Wang, "Sensing Terminal for Big Data Oil and Gas Pipeline Corrosion Monitoring," 2020 International Conference on Computer Network, Electronic and Automation (ICCNEA), 2020, pp. 279-283, doi: 10.1109/ICCNEA50255.2020.00064.

[9] T. Nguyen, R. G. Gosine and P. Warrian, "A Systematic Review of Big Data Analytics for Oil and Gas Industry 4.0," in IEEE Access, vol. 8, pp. 61183-61201, 2020, doi: 10.1109/ACCESS.2020.2979678.

[10] M. F. A. M. Fauzi, I. A. Aziz and A. Amiruddin, "The Prediction of Remaining Useful Life (RUL) in Oil and Gas Industry using Artificial Neural Network (ANN) Algorithm," 2019 IEEE Conference on Big Data and Analytics (ICBDA), 2019, pp. 7-11, doi: 10.1109/ICBDA47563.2019.8987015.

[11] M. S. A. Rahman, N. A. A. Jamaludin, Z. Zainol and T. M. T. Sembok, "Machine Learning Algorithm Model for Improving Business Decisions Making in Upstream Oil & Gas," 2021 International Congress of Advanced Technology and Engineering (ICOTEN), 2021, pp. 1-5, doi: 10.1109/ICOTEN52080.2021.9493499.

[12] A. Roubickova, N. Brown and O. Brown, "Using Machine Learning to Reduce Ensembles of Geological Models for Oil and Gas Exploration," 2019 IEEE/ACM 5th International Workshop on Data Analysis and Reduction for Big Scientific Data (DRBSD-5), 2019, pp. 42-49, doi: 10.1109/DRBSD-549595.2019.00012.

[13] A. Mohamed, M. S. Hamdi and S. Tahar, "A Machine Learning Approach for Big Data in Oil and Gas Pipelines," 2015 3rd International Conference on Future Internet of Things and Cloud, 2015, pp. 585-590, doi: 10.1109/FiCloud.2015.54.