Journal of Petroleum Research and Studies

Open Access
No. 34 part 2, April 2022, pp. 138-149

P- ISSN: 2220-5381
E- ISSN: 2710-1096

DOI: http://doi.org/10.52716/jprs.v12i1(Suppl.).627

Integrating CPI and Core Data into Logistic Regression for Lithofacies Modeling

Alaa M. Hasan[1*], Ahmed H. Sabe[2]

Basra Oil Company, Basra, Iraq

[*]Corresponding Author E-mail: alaamh1985@gmail.com

6th Iraq Oil and Gas Conference, 29-30/11/2021

Abstract

The lithofacies classification research is part of an extended multidisciplinary reservoir characterization and simulation study that has been implemented on the upper shale member/Zubair formation of the southern Iraqi X oil field. This study has been conducted through the Integrated Reservoir Management School (IRMS) at the Basrah Oil Company (BOC). Lithofacies classification is a process to determine rock lithology by analyzing core and well- log data set. Traditionally, lithofacies were classified manually or with the use of some graphing approaches. Many artificial intelligence techniques have recently been adopted to categorize lithofacies. In this work, two robust algorithms were applied to modeling the lithofacies through specific well section (formation), these procedures were adopted and their results were compared to determine more accurate lithofacies classification method. Logistic Boost Regression LBR and Multinomial Logistic Regression MLR were utilized to model the resulting lithofacies as a function of CPI dataset in order to anticipate discrete properties distribution in non-cored depth in wells.

CPI data, which are available for 49 wells in the upper shale Zubair formation, includes: water saturation, porosity (Ø_neutron) and volume of shale (V_sh). However, routine core analyses of permeability, porosity and facies are existent for only one well. For that well, the lithofacies types are sand, silty sand and shale. Two supervised statistical learning techniques, LBR and MLR, has been certified to model the discrete lithofacies distribution as a function of the CPI records. The lithofacies classification was then validated through forming the confusion table and computing the accuracy for each method. LBR was observed to be the optimum approach as it led to more accurate lithofacies classification than MLR in clastic reservoirs. The presented workflow demonstrated reasonable facies distribution that leads to strong relationship between porosity and permeability to estimate the petrophysical properties in non-cored wells.

In addition, the posterior lithofacies distribution were plotted to show the probability of spatial distribution and direction of model. These algorithms implemented through R programming a language commonly used in statistical computing by using software packages. Then, these costs for overall data process acquisition could be reduced.

Keywords: Geostatistical, classification, lithofacies, reservoir characterization, hydrocarbon, porosity, permeability.

دمج CPI والبيانات الأساسية في الانحدار اللوجستي لنمذجة الحصى

الخلاصة:

ان بحث تصنيف نوعية الصخور هو جزء من دراسة توصيف ومحاكاة المكمن والتي أجراها فريق متعددة التخصصات والتي تمت على حقل X/ تكوين الزبير والواقع في جنوب العراق. هذا البحث هو احدى نتاجات مدرسة الادارة المكمنية المتكاملة (IRMS) في شركة نفط البصرة .(BOC) ان تصنيف صخور الطبقات الجيولوجية تتم من خلال تحليل مجموعة البيانات الأساسية لباب الصخور core ومخططات مجسات الابار Logs.تقليديا ، تتم هذه العملية مختبرياً أو باستخدام بعض أساليب الرسوم البيانية. وفي الأونة الاخيرة تم اعتماد العديد من تقنيات الذكاء الاصطناعي لتصنيف الطبيعة الصخارية للطبقات. وفي هذا البحث ، تم تطبيق خوارزميتين موثوقتين LBR و MLR لنمذجة الطبيعة الصخارية المعلومة في الابار التي تحتوي على نماذج صخرية للطبقات الجيلوجية كدالة لمجموعة من تفسيرات المجسات CPI لنفس البئر وتطبيق هذا النموذج على بقية الابار التي لا تحتوي على نماذج لباب Core لنفس (التكوين).

ان البيانات المتاحة في هذه الدراسة تتضمن التسجيلات البئرية لـ 49 بئر لحقل الزبير/ تكوين الزبير / العطاء الاعلى والتي تحتوي على: التشبع المائي ، المسامية النيوترونية و حجم الصخوري. اما البيانات الخاصة بالتحاليل المختبرية لنماذج الصخور والتي يحسب منها: النفاذية والمسامية والسحنة فهي متاحة لبئر واحد فقط. ان التركيبة الصخرية الناتجة من التحليل المختبري لنموذج اللباب هو الرمل sand ، silty sand و shale.

تم اعتماد طريقتين من طرق الذكاء الاصطناعي وهي LBR : و MLR ، لنمذجة توزيع الصخور المنفصلة كدالة للتفسيرات البئريةCPI ، وتم التحقق من صحة عمل هاتين الطريقتين باستخدام جداول خاصة لحساب النسبة المئوية للخطأ، وقد لوحظ ان LBR هو النهج الأمثل لأنه أدى إلى تصنيف للطبيعة الصخرية أكثر دقة من MLR في المكامن الرملية.clastic

أظهر سير العمل المقدم توزيعاً معقولًا للسحنات يؤدي إلى علاقة قوية بين المسامية والنفاذية لتقدير الخصائص البتروفيزيائية في الآبار التي لا تحتوي على نماذج كور.

علاوة على ذلك ، تم رسم التوزيع الصخوري مع العمق لإظهار احتمالية التوزيع المكاني واتجاه النموذج. وقد تم تنفيذ هذه الخوارزميات من خلال برمجة R ، وهذه اللغة مجانية و شائعة الاستخدام في البرامجيات الاحصائية. وبأستخدام هذه الطريقة يمكن تخفيض تكاليف الحصول على البيانات المطلوبة.

1. Introduction

The quality of static and dynamic reservoir models depends on the rocks type and fluid properties inputs, such as lithofacies, porosity, hydrocarbon saturation of

Open Access
No. 34 part 2, April 2022, pp. 138-149

P- ISSN: 2220-5381
E- ISSN: 2710-1096

hydrocarbon fluids, and permeable of rocks [1]. Geo-statistical classification of lithofacies is a significant portion of reservoir description, which allows for the anticipation of rock facies types by using geo-statistical analysis of well- log data in depth of wells that have not been cored. Analyses and anticipation of lithology with fewer core data could decrease overall data acquisition costs and decrease overall project cycle time and budget for new field developments [2].

Adoption of rock facies classification into the modeling of formation permeability (K) and porosity ($\emptyset$), especially given the underlying measurement and interpretation of the well-log. This is a critical step to decrease uncertainty in reservoir description. [3]. Lithofacies (facies defined using core analysis) are a key property, since most other properties show response with facies variation. The discrete facies sequence is produced either from core analysis in the laboratory (lithofacies) or adopted cluster method to grouping logging data and informed electrofacies [4].

As mentioned above, facies classification carried out in specifying rock sorts or classes to a particular sample based on measured features. This process considered important to boosting the relationship between measured permeability K and porosity $\emptyset$ and their results affectively utilize to predicting the other petro-physical properties in non-cored depths or wells. Moreover, lithofacies classification of rocks is a crucial step in structural interpretation of seismic because different lithofacies have seen variations in permeability K and rock saturation by water Sw for a measured porosity $\emptyset$ [5]. Therefore, several alternative methods to find facies types from well-log dataset have been suggested, like classical multivariate statistical approaches and neural networks [6].

This work presents two geo-statistical techniques for identifying and specifying lithofacies through utilizing Logistic Boost Regression (LBR) and Multinomial Logistic Regression (MLR) to model the discrete properties (lithofacies) as a function of available well-logs interpretation CPI from the logging dataset.

2. Field And Reservoir Description

Classification and prediction were performed on the X oil field dataset for a reservoir with sandstone, shale and siltysand layers as well limestone at lower and

Open Access
No. 34 part 2, April 2022, pp. 138-149

P- ISSN: 2220-5381
E- ISSN: 2710-1096

above parts of the reservoir. Zubair formation is divided into three members, which are upper shale, sand and lower shale.

X oil field is one of Basrah oilfields in the south of Iraq about 100 km north-west of the center of Basrah city and 80 km north of Rumelia oilfield as shown in Figure (1). The API gravity for produced oil reached 32^o [7].

3. Research Methodology

The classification represents modeling the observed discrete lithofacies in a well depth as a function of well-logging interpretations of CPI for the purpose of estimating their distribution in missing depths in the same well or other wells that have unknown facies [8]. In this paper, the Logistic Boosting Regression (LogitBoost) was utilized as an effective classification approach to model the measured discrete distribution of lithofacies in the well of the sandstone reservoir under study.

3.1. Logistic Boosting Regression (LBR)

Logistic Boosting Regression is one of an efficient computational in classification methodology for learning accurate classifiers. this approach starts by sequentially applying a classification algorithm to reweighted blocks of training dataset and then taking a weighted majority vote of the series of classifiers that produced. Recently this algorithm utilized widely to a variety of classification problems, due to its ease of implementation, show high accuracy, and don't need an external tool to optimize the algorithm [9].

This algorithm (LogitBoost), Observable as a multilateral, measured regression function from the statistical aspect. For each time, we can fit individually weighted function regression (base learner) to recognize the optimization progressively [9]. An implementation of extensions to Freund, Schapiro's AdaBoost Algorithm, and Friedman's gradient boosting machine. Includes regression functions for least squares, absolute loss, t-distribution loss, quantile regression, logistic, and multinomial logistic [10].

Open Access
No. 34 part 2, April 2022, pp. 138-149

P- ISSN: 2220-5381
E- ISSN: 2710-1096

3.2.Multinomial Logistic Regression (MLR)

This algorithm is a classification approach that generalizes logistic regression to multi-set categories, i.e. with more than two possible discrete results. This sort is utilized to estimate categorical placement in, or the probability of category membership on a dependent variable based on multiple independent variables. This variable can be either binary, interval or ratio in scale.

Multinomial logistic regression is an easy expansion of (binary logistic regression) that gives for more than two groups of the dependent or result variable. Like (binary logistic regression), (multinomial logistic regression) utilizes maximum likelihood predict to estimate the probability of categorical membership [11].

(Multinomial logistic regression) require accurate regard of the sample size and check for isolated states. Like other dataset analysis proceedings, initial data analysis should be thorough and include careful univariate, bivariate, and multivariate assessment.

precisely, multi-collinearity should be estimated with simple regression through independent variables. Also, standard multiple regression (multivariate diagnostics) can be utilized to reach for multivariate outliers and for the exclusion of outliers or influential cases [2].

4. Results and Discussion

In this paper, two algorithms were applied for lithofacies classification on the data of X oil field. Firstly, the modeling of classification procedure start with correlated discrete properties (lithofacies) with well-log interpretation CPI (i.e. porosity $\emptyset$, water saturation Sw and volume of shale Vsh). Fig.2 & Fig.3 shows the relationship between each two variables in the well-logs & core dataset: depth, core porosity, core permeability, saturation of water, shale of volume and log porosity. The sand, siltysand and shale were shown in Fig.4 to represent the properties variation for each facies. The X dataset display and summarize as shown below:

Journal of Petroleum Research and Studies

Open Access
No. 34 part 2, April 2022, pp. 138-149

P- ISSN: 2220-5381
E- ISSN: 2710-1096

```
> head(data)
  Depth.m Core.Porosity Core.Permeabilty   SW  Vsh Porosity.Log LithoFacies
1 2786.66         0.031             0.19 0.99 0.24         0.00  silty sand
2 2789.40         0.035             0.19 0.98 0.51         0.03  silty sand
3 2789.86         0.028             0.19 0.98 0.35         0.01  silty sand
4 2790.35         0.074             0.19 0.50 0.43         0.07  silty sand
5 2790.77         0.124             0.75 0.99 0.59         0.05  silty sand
6 2794.26         0.068             0.68 0.99 0.67         0.00  silty sand
```

```
> summary(data)
    Depth.m       Core.Porosity    Core.Permeabilty        SW
 Min.   :2787   Min.   :0.0280   Min.   :   0.15   Min.   :0.0500
 1st Qu.:2804   1st Qu.:0.1270   1st Qu.:  10.45   1st Qu.:0.1700
 Median :2814   Median :0.2000   Median : 168.00   Median :0.5400
 Mean   :2815   Mean   :0.1796   Mean   : 422.09   Mean   :0.5667
 3rd Qu.:2827   3rd Qu.:0.2380   3rd Qu.: 707.00   3rd Qu.:0.9900
 Max.   :2839   Max.   :0.2620   Max.   :4186.00   Max.   :0.9900
      Vsh          Porosity.Log        LithoFacies
 Min.   :0.0200   Min.   :0.0000   sand      :59
 1st Qu.:0.0500   1st Qu.:0.0600   shale     : 1
 Median :0.1200   Median :0.1500   silty sand:21
 Mean   :0.1921   Mean   :0.1322
 3rd Qu.:0.2800   3rd Qu.:0.2100
 Max.   :0.6700   Max.   :0.2400
```

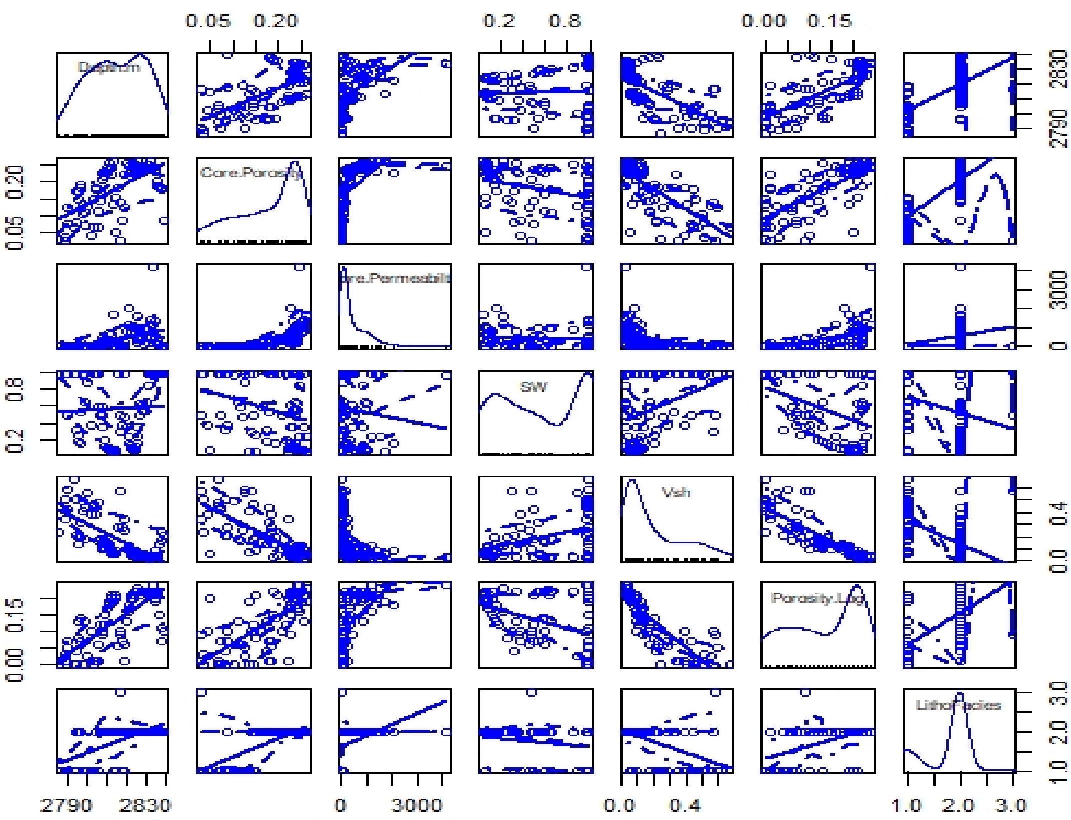

Fig. (2): The plots are shown the relationship between core & well-logs

Open Access
No. 34 part 2, April 2022, pp. 138-149

P- ISSN: 2220-5381
E- ISSN: 2710-1096

CPI variables as Scatter matrix

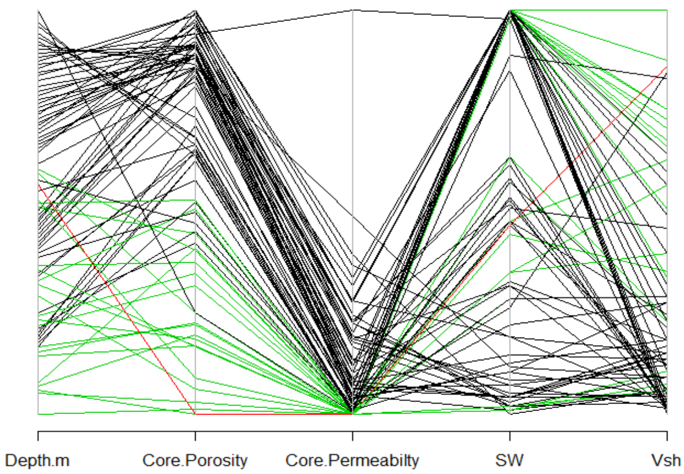

Fig. (3): Variations for CPI well-logs & core properties for each lithofacies

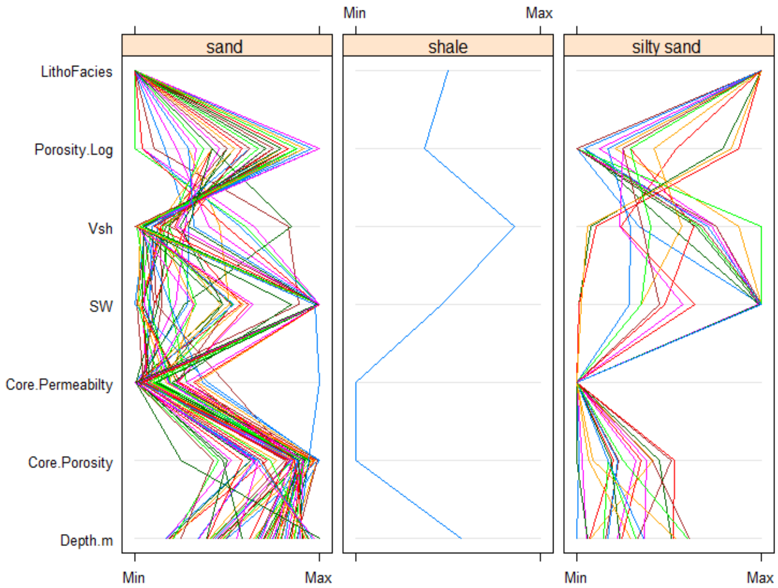

Fig. (4): Variation CPI well-logs & core properties as a Parallel plot for each Lithofacies

The accuracy of classification approaches was measured based on accuracy of the estimated discrete lithofacies distribution. In this work, we find posterior probability distribution of rock facies was conducted through MLR & LBR as shown in Figures (5) & (6).

Open Access
No. 34 part 2, April 2022, pp. 138-149

P- ISSN: 2220-5381
E- ISSN: 2710-1096

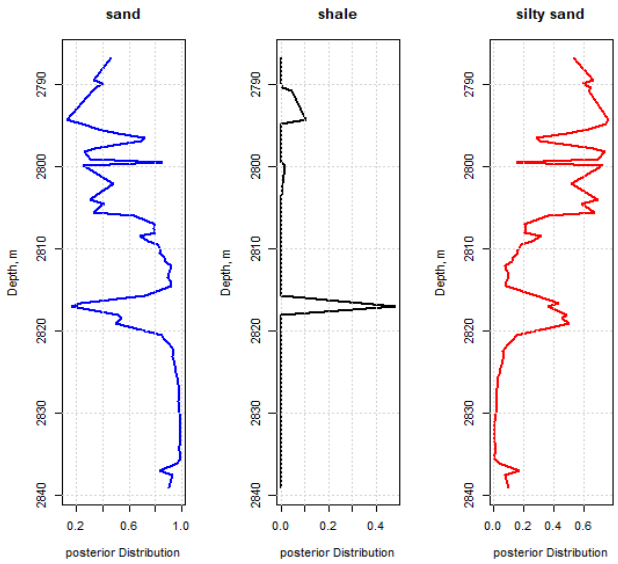

Fig. (5): Total percentage for each estimated facies by MLR along with the Well depths.

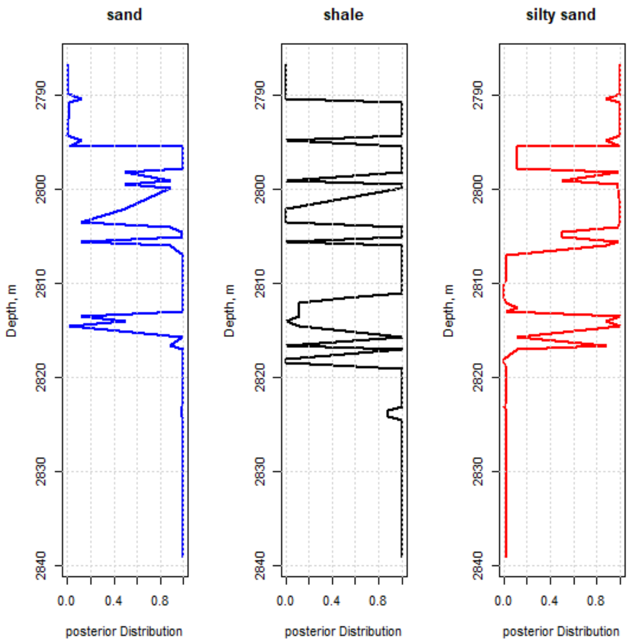

Fig. (6): Total percentage for each estimated facies by LBR along with the Well depths.

The total percent correct of the estimated discrete properties (lithfacies) for Logistic Boost Regression and Multinomial Logistic Regression were 100% and 80%, respectively.

For further illustration about the efficiency of logistic boosting regression and

Journal of Petroleum Research and Studies

Open Access
No. 34 part 2, April 2022, pp. 138-149

P- ISSN: 2220-5381
E- ISSN: 2710-1096

Multinomial Logistic Regression lithofacies classification approaches, the predicted discrete lithofacies given the test subset were decorating in a boxplot matching with the observed lithofacies, as shown in Figure (7).

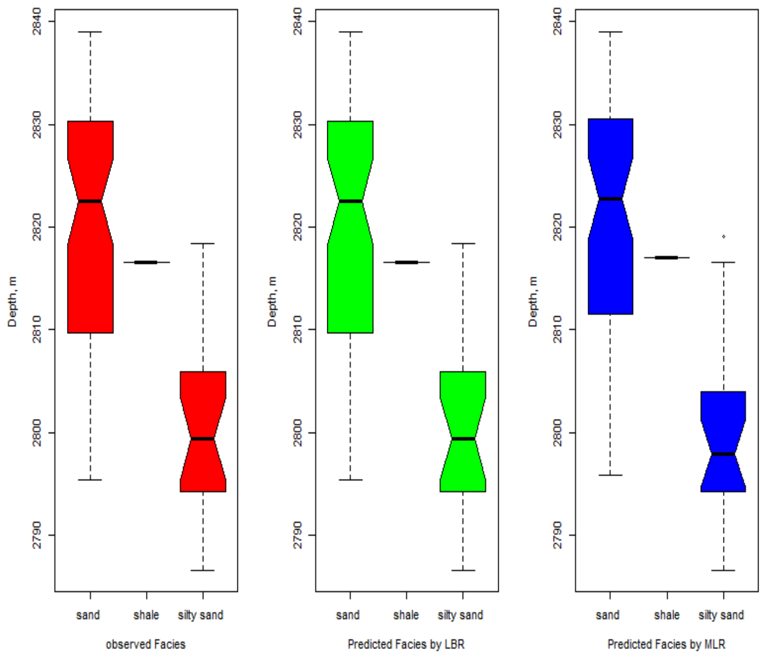

Fig. (7): Compare the real lithofacies and predicted facies with depths by utilized LBR & MLR approach.

The summations of diagonal table for predict and observed Lithofacies for LBR:

```
             sand shale silty sand
sand           59     0          0
shale           0     1          0
silty sand      0     0         21

> diag(prop.table(ct2))
      sand      shale silty sand
0.72839506 0.01234568 0.25925926

> sum(diag(prop.table(ct2)))
[1] 1
```

The summations of diagonal table for predict and observed Lithofacies for MLR:

Open Access
No. 34 part 2, April 2022, pp. 138-149

P- ISSN: 2220-5381
E- ISSN: 2710-1096

```
             sand shale silty sand
sand           51     1          7
shale           0     0          1
silty sand      7     0         14

> diag(prop.table(ct4))
      sand      shale silty sand
 0.6296296  0.0000000  0.1728395

> sum(diag(prop.table(ct4)))
[1] 0.8024691
```

5. Conclusions:

Two statistical algorithms were taken on models for the observed discrete properties (lithofacies) correlated with core and well-log dataset. The two algorithms of facies classification include: Logistic Boosting Regression and Multinomial Logistic Regression. The comparison between these algorithms were the total correct percent of the different facies in (The summations of diagonal table for predict and observed Lithofacies). This case study observed that Logistic Boosting Regression is the most accurate modeling as resulted in 100% total correct percent of facies classification, Multinomial Logistic Regression resulted 80% total correct percent of facies classification.

Open Access
No. 34 part 2, April 2022, pp. 138-149

P- ISSN: 2220-5381
E- ISSN: 2710-1096

References:

[1] McCreery, E. B., and W. J. Al-Mudhafar. "Geostatistical Classification of Lithology Using Partitioning Algorithms on Well Log Data-A Case Study in Forest Hill Oil Field, East Texas Basin." *79th EAGE Conference and Exhibition 2017*. European Association of Geoscientists & Engineers, 2017. https://doi.org/10.3997/2214-4609.201700905

[2] Al-Mudhafar, W. J. Advanced Supervised Machine Learning Algorithms for Efficient Electrofacies Classification of a Carbonate Reservoir In a Giant Southern Iraqi Oil Field. *Offshore Technology Conference, Houston*, TX, USA. 2020, https://doi.org/10.4043/ 30906-MS

[3] Al-Mudhafar, Watheq J. "Applied Geostatistical Reservoir Characterization in R: Review and Implementation of Permeability Estimation Modeling and Prediction Algorithms-Part II." *Offshore Technology Conference*., *Houston*, TX, USA. 2016. https://doi.org/10.4043/ 26932-MS

[4] Schafer, J. L. Analysis of incomplete multivariate data. Chapman and Hall/CRC. 1997.

[5] Rogers, F. J., & Iglesias, C. A. Rosseland mean opacities for variable compositions. *The Astrophysical Journal*, no. 401, pp. 361-366 . USA. 1992.

[6] Zhao, T., Jayaram, V., Roy, A., & Marfurt, K. J. "A comparison of classification techniques for seismic facies recognition". *Interpretation*, no. 3, pp. 29-58. SEG Library. 2015. https://doi.org/10.1190/ INT-2015-0044.1

[7] Al-Mudhafar, W. J. "Integrating well log interpretations for lithofacies classification and permeability modeling through advanced machine learning algorithms". *Journal of Petroleum Exploration and Production Technology*, no. 7, pp. 1023-1033. 2017. https:// doi.org/10.1007/s13202-

Open Access
No. 34 part 2, April 2022, pp. 138-149

P- ISSN: 2220-5381
E- ISSN: 2710-1096

017-0360-0

[8] A. Alher, M. Aljawad, and A. Ali, “Static Model of Zubair Reservoir in X Oil Field”, *ijcpe*, vol. 19, no. 1, pp. 57-60, Mar. 2018.

[9] Friedman, J., Hastie, T., & Tibshirani, R. “Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)”. The annals of statistics, no. 28, pp. 337-407. April 2000, https://doi.org/10.1214/aos/1016218223

[10] Xing, C., Geng, X., & Xue, H. Logistic boosting regression for label distribution learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4489-4497. 2016

[11] Rogers, S. J., Fang, J. H., Karr, C. L., & Stanley, D. A. (1992). Determination of lithology from well logs using a neural network. *AAPG bulletin*, vol.76, no.5, pp. 731-739. 1992, https://doi.org/10.1306/BDFF88BC-1718-11D7-8645000102C1865D

[12] Dixit, S., Kumar, B., Singh, A., & Ashoka, R. An application of multinomial logistic regression to assess the factors affecting the women to be underweight and overweight: a practical approach. *Intl J Health Sci Res*, no. 5, pp. 11-17. 2015