Volume 5, No. 8, Nov-Dec 2014

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Effectiveness Logistic Regression Approach for Petroleum Production Ratio

Nagwa.L.Badr
Ain Shams University: Information System department
Faculty of Computer and Information Sciences
Cairo, Egypt

Abstract: An oil reservoir has become one of the most important areas in the world on economic and environmental sustainability. Intelligent petroleum systems concerns of development oil industry. The interval between intelligent systems in petroleum domain is the ratio of results accuracy. Using regression model seek for systems gaps. Regression model involves a re-examination of the system results based on failures proportion and fixed faults that have re-emerged. In the meantime, re-exam results achieved percentage accuracy in the system and the efficiency. In this paper, re-test the standard petroleum factors values through a regression model to enhanced oil production. Evaluate the accurately results contribute for enhance different fields on oil industry domain.

Keywords*:* Intelligent System; Regression; Test Cases; Petroleum; Prediction

I. INTRODUCTION

One of the programming software development integral parts is the regression [1]. Consequently, regression testing has traditionally been performed by a software quality assurance team after the development team has completed work, grey system for Iranian production ratio is a sample [2]. Conceptually, this problem is being addressed by the rise of system re-testing. Development cycle in regression model depends on the test cases; re-run process has been generally either functional tests or unit tests to outcomes verification [3]. Developer testing compels a developer to focus on unit testing and to include both positive and negative test cases [4]. Production with regression model is considered, based on data which was collected from different sources.

The previous systems values shows the environmental factors affect the model and what is the importance of environmental factors. Use regression test functions try to predict dependent variables as a function of other correlated observable independent variables [5]. As a remedy, regressions approve to re-test system values and analysis the obtaining forecasts [6]. Regression techniques have long been central to the World of economic statistics (“econometrics”). Increasingly, they have become important on different domains such as lawyers and legal policy makers as well [7]. Among most methods which are used in crude oil production are tested through a regression model. In petroleum systems; using regression model to prove values enhanced oil industry; multiple characteristics on oil domain such as rock porosity and permeability [8].

Over the life of the reservoir, many crucial decisions depend on formation estimates [9]. Furthermore, the measurement site in petroleum domain is available and limited to isolate well locations [10]. A regression model was made to show the effect of re-testing the prediction processes of petroleum factors. The evaluation study applies test cases, regression models and time series forecasting of vague petroleum datasets to achieve more accurate results. The paper goal is to build complete petroleum intelligent system that contains different functions such as: classification, data mining, prediction, test cases; regression and time series forecasting models.

The rest of the paper is organized through different sections: Section 2 shows related work. Section 3, explains regression approach; Section 4 shows system architecture. In section 5, scenario experiments of the model and finally, Section 5, shows the experimental scenario. In section 6 evaluation processes are explained whereas in section 7, the paper conclusion and future work suggestions are discussed.

II. RELATED WORK

Most related work on regression testing has used coverage information. The petroleum domain contain huge amount of data, geophysics, images, remote sensors sources and other. Using regressing model provides test cases in terms of sequence base on the number of impacted crude oil dataset; these dataset are collected from distinct sources [11].

Using regression test within static coverage values for prediction purpose is achieved with multiple test cases [12]. The proposed regression test neglects the petroleum software code, previous predicted results and historical values. Interesting about retesting the test cases statues and prove regression test of the system results. Another related study is the Binary Matching Tool (BMAT), which compute the changes between two versions of the program at the basic block granularity. It represents another proposed a conceptual way to combine minimization and prioritization to select test cases. It suggested that prioritization could have a time complexity [13]. An algorithm with general regression neural network to build oil prices forecasting model is discussed [14]. Another proposed research is a history-based test technique. It is display a Reliable model for estimating the Wax deposition rate during crude oil production and processing. One main assumption of this technique is that historical test case performance data might be used to improve long run regression testing performance [15].

On the other hand, impacts of unconventional gas development on China's natural gas production are approved. It shows that the development of oil and gas ratio depends on more accurate results and highly precision value [16]. A prioritized list of test cases in regression model that predict for failures is another research case [17].

Systems improvements represents "re-run" process of outputs compared against expected results [18]. Simultaneously, the regression test concerns of failures of software's code, multiple projects are interest about retesting the software failures such as PETS project (Prediction of software Error rates based on test and Software maturity results) funded by the European Commission [19].

III. REGRESSION APPROACH

Regression analysis defines as statistical tool that aims the investigation within different variables. Conceptually, regression techniques have long been investigate hypothesis, imagine that gathering data on multiple domain. In petroleum domain, strictness prediction system has been approved and produced accurate predicted results [20]. Development of the oilfield production is the basis of the optimal decision making of oilfield manager. Furthermore, there are many methods to enhance the output of oilfield such as multiple linear regression, artificial neural network, grey prediction method, logistic curve method and other [21]. Realistically, the results accuracy of the intelligent systems never achieves a complete precision value. Test case processes and regression test aims to validate system failures and focus on real factor which affect of results. In this case, test case process applying on the petroleum prediction system to emphasis the predicted result; regression test reduce the test case number, minimize error time and declares more precision of the system. Material in this work is an extension of statistical regression analysis with normal distributed data and mean=$\beta0+\beta1\chi1+\cdots+\beta sk$

Where Xi represents explanatory variables acquitted from distinct resources. Retest the dataset depends of the following reasons [22]:

A. Data not necessarily from a normal distribution.
B. Data may be censored.
C. Non standard regression models that relate life to explanatory variables.
D. Presentation motivated by practical problems in reliability analysis.

Different petroleum reservoirs variables are retested through studding every membership separately and show which the most affecting of the production results. Figure (1), display sample of reservoirs petroleum memberships which are affecting of petroleum production.

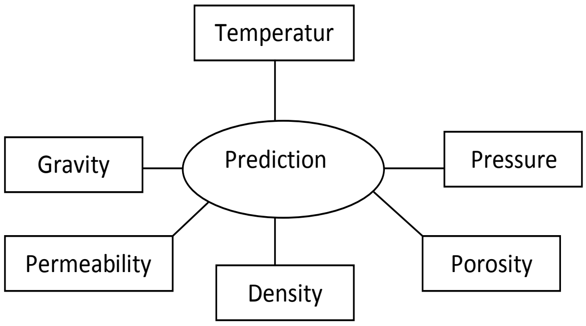

Figure 1. Reservoirs memberships samples

Figure (2), explains different intelligent systems which are concerns of petroleum production domain, regression model tools and the techniques that used in the same domain.

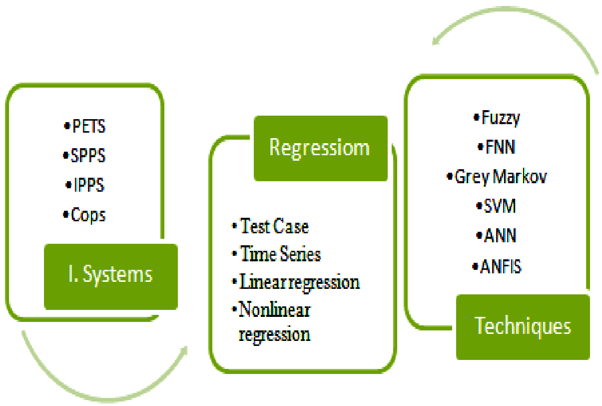

Figure 2. Regression & intelligent Systems

IV. SYSTEM ARCHITECTURE

Different petroleum reservoirs variables are retested through studding every membership separately and show which the most affecting of the production results. Figure (3), display sample of reservoirs petroleum memberships which are affecting of petroleum production. In addition, the system architecture consists of two basic modules:

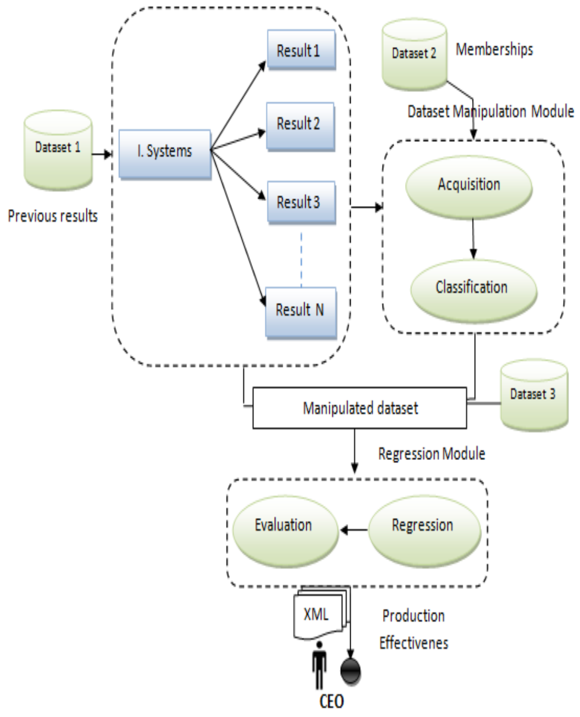

Figure 3. System Architecture

A. Dataset acquisition Module: One of the system requirements is the data; dataset acquisition is a module, the major purpose is to upload and reformat the dataset, which is collected from different sources into two parts: Dynamic and static dataset acquisition. Multiple petroleum intelligent systems production results are collected as a discrete dataset, membership functions are collected into another data base. The dynamic dataset has the ability to collect from experts, whereas the static dataset includes journals, web sites, oilfields and other sources. During data sorting, the modules functions concerns of dataset classifications by using Weka [23]. Consequently, the standards petroleum properties which are approved are exploited to re-testing process through a regression model that prove membership's production effectiveness based on dataset classification.

Regression Module: regression model functions exploiting the classified dataset which are manipulated to re-run the effectiveness of crude oil properties into production process. On another hand, apply regression algorithm detects the membership which are affecting of crude oil production. Using data mining tools aims to refine the factors which highlighting the effectiveness factors on production process. Regression model represents the technique that ensures accurately systems values and precision ratio.

V. REGRESSION EXPERIMENTAL SCENARIO

At present time, regression model is one sample of several major models which are being used to develop the production processes in the oilfield. The problem is, there are several input variables in the above models and significant factors influencing intelligent system is not considered. Thus, the result is not accurate. Meanwhile, the regression model is more simple and accurate. Table (1), shows the standard petroleum memberships values which are used through different intelligent systems and applications, as follows:

M#	*Temperature*	*Pressure*	*Density*	*Gravity*	*Gas-Density*
1	108.3361	67.59	0.861	32.7	0.8528
2	62.8573	71.523	0.851	34.7	0.8529
3	57.2615	55.684	0.871	30.9	0.8603
4	85.0587	97.034	0.871	30.9	0.8603
5	85.6211	40.454	0.875	30.1	0.8592

There are more than thirty three memberships which are effects of petroleum production process. Most of the intelligent systems in petroleum domain used the standard factors. Five petroleum parameters (Temperature, pressure, gravity, density, permeability, porosity and gas-density) are tested. The effectiveness dataset of petroleum factors shown on figure (4), where x- axis represent petroleum factors sample whereas Y-axis explain probability effectiveness percentage, as follow:

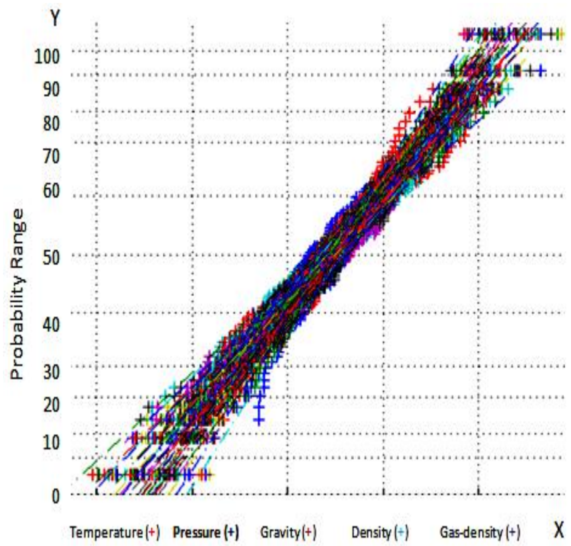

Figure 4. Factors Effectiveness probabilities

In the regression model different algorithms has ability to use. The logistic regression of factors as test cases calculated through Logit p(x) function (2) where x represents parameters values, β is the intercept values and X is test case parameter, sample is shown on table (2) [10]:
logiyP(X)=|n P(X)/(1-P(X))

Where

P(x) = 1/ [1+exp (β0+β1Xi)] (2)

The major goal of this model is to retesting the statues values through regression model. The initial regression shows the relationship between one independent variable (X) and a dependent variable (Y):

Y=β0+β1X+u (3)

Where the value of intercept represents by (β0), the value of slope represents as (β1) and (u) is the no predicted variation by the slope and intercept terms. Conceptually, slope and intercept represents through:

$$\text{Slope (m)} = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x^2)} \quad (4)$$

$$\text{Intercept} = \frac{n(\sum y) - m(\sum x)}{n} \quad (5)$$

Where x, y is the relational cases, n is the number of cases. Consequently, regression analysis for petroleum test cases data which are collected from distinct sources calculated through the following functions:

$$\bar{x}, \bar{y}, S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{[\sum_{i=1}^{n} x_i]^2}{n} \quad (6)$$

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n} \quad (7)$$

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}, S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n} \quad (8)$$

$$b_0 = \bar{y} - b_1 \bar{x}, \quad b_1 = \frac{S_{xy}}{S_{xx}} \quad (9)$$

$$S_{yy}/x = S_{yy} - \frac{(S_{xy})^2}{S_{xx}} \quad (10)$$

R= Logit P(X) * β- ɳ (11)

Where X1-Xn, Y1-Yn are the input data; S is the summation operation; b0, b1 are the slope, intercept estimate of regression curve and n is the number of experimental datasets.
Furthermore, using R function as a linear regression:

Where ɳ is the factor usage number and β is the correlation coefficient between distinct parameter values, shown in table (2):

Regression #	Temperature	Pressure	Density	Gravity	Gas-Density
1	108.3	67.5	0.8619	32.7	0.8528
2	62.8	71.5	0.8514	34.7	0.8529
3	57.2	55.6	0.8715	30.9	0.8603
4	85.0	97.0	0.8715	30.9	0.8603
5	85.6	40.4	0.8754	30.1	0.8592
R (%)	**78**	**87.2**	**54.9**	**64**	**67**

Using normal probability function norm(data), depends on regression values of the petroleum factors and previous system values; it shows that the pressure factor has more affective of oil and gas production within 87.2%. As a remedy of logistic regression model, the calculations of P(X) for the five factors in the regression model, it shows the following:

- Pressure values are the highest value.
- Temperature is the second parameter affect production process.
- Reduce the number of test cases aims to achieve more accurate production results.
- Minimizing of regression time error depends on classifying inputs and split test cases into several classes and subclasses.

VI. EVALUATION

Obviously, in petroleum domain; the highly cost of drilling process ensure that the validation of the predicted result has been priority. It donates the results more precision and aims to support Chief Executive Officer (CEO) to drilling or nullity. Using precision algorithm function [25], as follows:

$$Precision\,(x) = \frac{Num\ of\ factors\ (\beta)}{Num\ of\ predicted\ wells} \quad (12)$$

Where n represents the number of predicted oil wells and β is a positive real weight. In the precision function (4), through thirty predicted well, the precision of factors shows that the pressure factor represents the highly value 0.87; whereas density factor is the lowest one.
Through comparing previous systems and empirical values, it achieve that the pressure factor represent the highest value

on petroleum production process. In figure (5), the pressure factor represent the highest value which effect of the production process in different stages. X-axis represents the tested wells number. On another hand, Y-axis shows the different petroleum factors range builds on regression model values; as follows:

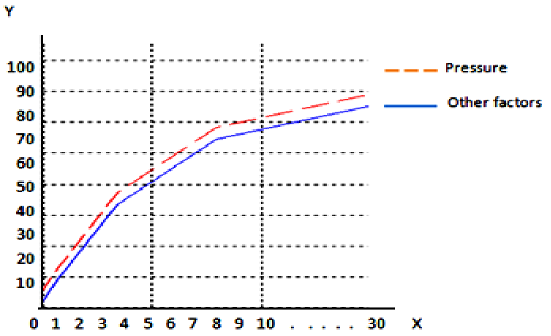

Figure 5. Intelligent system & memberships

VII. CONCLUSION

Development petroleum is a wide field; meanwhile, building knowledge base for oil domain aim experts and engineering's to enhance oil industries. More than one source of petroleum dataset is exploited. Review previously system values of production rate and classification data using Weka tool are discussed. This work applied through logistic regression model that concern of petroleum membership's effectiveness of production values. The model re-test previous system results and re-run petroleum factors. The results show that the pressure factor represents the highest values as 87.2%; whereas density factor is the lowest one. Evaluation processes ensure that factor close to the empirical production ratio. Hope in future work to testing more crude oil properties that enhances production process.

VIII. REFERENCES

[1] G. Rothermel, R. H. Untch, C. Chu, and M. J. Harrold," Prioritizing Test Cases For Regression Testing", in Software Engineering, IEEE Transactions, Vol. 27,2002, p. 929 – 948, doi: 10.1145/347636.348910.

[2] Jung-Min Kim, Adam Porter,"A history-based test prioritization technique for regression testing in resource constrained environments", in ICSE, May, 2002, pp: 119-129 ,doi: 1145/581339.581357.

[3] E Ranaee, G Porta, M Riva, and A Guadagnini.," Investigation of Saturation Dependency of Oil Relative Permeability During Wag Process through Linear and Non-Linear Pca". ECMOR XIV-14th European conference on the mathematic of oil recovery, 2004. doi: 10.3997/2214-4609.20141800

[4] Guangyu Dong, Jia Xu, Yuanyuan Song," Static Coverage Prediction for Regression Test, in Software Engineering", IEEE Transactions, Vol. 27,2004 pp. 248-236,doi: 10.4236/jsea.2014.78057.

[5] Senan A.Ghallab, Nagwa Badr, Abdel Badeeh Salem and Mohamed Fahmy Tolba," Logistic Test Case Regression Model of Petroleum Availability Prediction System", Advanced Machine Learning Technologies and Applications, Vol.488, 2014, pp. 248-257, doi:10.1007/978-3-319-13461-1.

[6] Senan A.Ghallab, Nagwa Badr, Abdel Badeeh Salem and Mohamed FahmyTolba, "A Fuzzy Expert System For Petroleum Prediction", WSEAS, 2013,p. 77-82, doi: 10.1016/j.jfa.2010.05.020

[7] Senan A Ghallab, NL Badr, and Mohamed F Tolba.," Intelligence Test Case Based-Approach for Crude Oil Prediction System'", in Intelligent Systems' 14,2015, p. 893-903 Springer, doi: 10.1007/978-3-319-11310-4_78

[8] Yanbin Li, Jiuju Zhang, Junming Xiao, and Yang Tan," Short-Term Prediction of the Output Power of Pv System Based on Improved Grey Prediction Model", in Advanced Mechatronic Systems (ICAMechS), International Conference on(IEEE),2014,pp:547-51,doi: 10.1109/ICAMechS.

[9] Wen-Tsao Pan, "Mixed Modified Fruit Fly Optimization Algorithm with General Regression Neural Network to Build Oil Prices Forecasting Model. Kybernetes", Vol. 43,2014, pp:53-63, doi: 10.1108/K-02-2014-0024.

[10] Kamari, Mohammadi, A Bahadori, and S Zendehboudi," A Reliable Model for Estimating the Wax Deposition Rate During Crude Oil Production and Processing", Petroleum Science and Technology, Vol.32, 2014, pp. 37-44,doi: 10.1080/10916466.2014.919007.

[11] Ting Wang, and Boqiang Lin,"Impacts of Unconventional Gas Development on China 'S Natural Gas Production and Import", Renewable and Sustainable Energy, Vol. 39, 2014, pp. 46-45, doi: 10.1016/j.rser.2014.07.

[12] M. Mayo and S. Spacey," Predicting Regression Test Failures using Genetic Algorithm-Selected Dynamic Performance Analysis Metrics", SSBSE 2013, pp. 158-171, doi: 10.1049/ip-sen:20030559.

[13] R Ahmadov, S McKeen, M Trainer, R Banta, A Brewer, S Brown, PM Edwards, JA de Gouw, GJ Frost, and J Gilman, "Understanding High Wintertime Ozone Pollution Events in an Oil and Natural Gas Producing Region of the Western Us", Atmospheric Chemistry and Physics Discussions", Vol.14, 2014, pp. 295-343, doi: 10.5194/acpd-14-20295-2014.

[14] Kordnoori, Mostafaei, "Grey markov model for predicting the crude oil production and exportat in Iran", International Journal of Academic Research, Vol.3 (2), 2013, pp. 97-102. Doi: dx.doi.org/10.7813/2075-4124.

[15] M. Mayo and S .Space, "Predicting Regression Test Failures using Genetic Algorithm-Selected Dynamic Performance Analysis Metrics", Technical report, New Zeland university, 2004. doi.org/10.1007/978-3-642-39742-4_13.

[16] Ahmed M. Salem ,Kamel Rekab ,James A. Whittaker," Prediction of software failures through logistic regression", Information and Software Technology, Elsevier press, 2004, pp. 781-789,doi:

[17] Fatai Adesina Anifowose Abdulazeez Abdulraheem." A Functional Networks-Type-2 Fuzzy Logic Hybrid Model for the Prediction of Porosity and Permeability of Oil and Gas Reservoirs",Proceeding of Computational Intelligence", Modelling and Simulation , pp. 193-198, 2009.doi:

[18] Michael Grottke, Klaudia Dussa-Zieger," Prediction of Software Failures Based on Systematic Testing", In proceeding of (EuroSTAR), 2009, pp. 1-12, doi: 10.1145/2593833.2593842.

[19] Dun Liu, Tianrui Li, and Decui Liang, "Incorporating Logistic Regression to Decision-Theoretic Rough Sets for Classifications", Approximate Reasoning", Vol. 55, 2014, pp. 75-87, doi: 10.1016/j.ijar.2013.02.013.

[20] Ahmed Senouci, Mohamed Elabbasy, Emad Elwakil, Bassem Abdrabou, and Tarek Zayed," A Model for Predicting Failure of Oil Pipelines". Structure and Infrastructure Engineering, Vol.10, 2013, pp. 375-87 , doi: 10.1080/15732479.2012.756918.

[21] JM Hahne, F Biebmann, N Jiang, H Rehbaum, D Farina, FC Meinecke, K-R Muller, and LC Parra," Linear and Nonlinear Regression Techniques for Simultaneous and Proportional Myoelectric Control". Neural Systems and Rehabilitation Engineering, IEEE Transactions on,Vol. 22 ,2014, pp. 69-79, doi: 10.1109/TNSRE.2014.2305520.

[22] Mohammed S El-Abbasy, Ahmed Senouci, Tarek Zayed, Farid Mirahadi, and Laya Parvizsedghy," Artificial Neural Network Models for Predicting Condition of Offshore Oil and Gas Pipelines", Automation in Construction, Vol.45,2014,pp. 50-65, doi: 10.1061/(ASCE)CO.1943-7862.0000838.

[23] MA Lin, DING Yong," The Data Mining Research of Library Which Base on Weka", Computer Knowledge and Technology, Vol. 24, pp. 149-153, 2009, doi: 10.1145/1656274.1656278.

[24] Enrique L Droguett, Isis D Lins, Márcio C Moura, Enrico Zio, and Carlos M Jacinto, 'Variable Selection and Uncertainty Analysis of Scale Growth Rate under Pre-Salt Oil Wells Conditions Using Support Vector Regression", Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability , 2014.doi: 10.1177/1748006X14533105.

[25] D.M. Powers," A computationally and cognitively plausible model of supervised and unsupervised learning", Advances in Brain Inspired Cognitive Systems, Vol. 119 (7), pp. 145-156 ,2013,doi: 10.1007/978-3-642-38786-9_17.

