Applications of Artificial Intelligence (AI) in Petroleum Engineering Problems

Shan e Zehra Lashari

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Recommended Citation
https://researchrepository.wvu.edu/etd/6041

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.
Applications of Artificial Intelligence (AI) in Petroleum Engineering Problems

Shan e Zehra Lashari

Thesis submitted
To the Benjamin M. Statler College of Engineering and Mineral Resources
at West Virginia University

in partial fulfillment of the requirements for the degree of

Master of Sciences in
Petroleum and Natural Gas Engineering

Ali Tadbiri Borujeni, PhD., Chair
Ebrahim Fathi, PhD.
Fatemeh Belyadi, PhD.
Ming Gu, PhD.

Department of Petroleum and Natural Gas Engineering

Morgantown, West Virginia
2018

Keywords: AI, Regression, Clustering, Classification, Machine Learning, ROP, Well Logs, Fuzzy logic, Smart Proxy, Reservoir Fluid Characterization

Copyright 2018 Shan-E Lashari
Abstract

Applications of Artificial Intelligence (AI) in Petroleum Engineering Problems

Shan e Zehra Lashari

For the last few decades multiple business sectors have been influenced by the advancement in Artificial Intelligence (AI). Though the oil and gas sector began to utilize the potential of AI comparatively latter than many other sectors, the appreciable amount of work has been done by researchers to equip the industry with AI tools. This work aims to explore various horizons of petroleum engineering by using different AI tools.

For providing better decision making in reservoir fluid characterization problem, fuzzy logic has been applied, which is an AI method to drive decisions when data is incomplete or unreliable. The second part of the work is the combination of supervised and unsupervised machine learning has provided an automated version of well log analysis, where the generated algorithm is able to distinguish between different lithological zones on the basis of well log parameters.

The majority of the problems such as drilling process optimization, production forecasting, comes under the umbrella of statistical regression. The supervised learning regression algorithm was generated to predict the drilling performance in terms of rate of penetration. The similar model was used for producing regression analysis of reservoir that has been treated by steam assisted gas drainage. The accuracy of both cases were investigated by comparing the prediction with available real time data.

The work has been concluded by providing conclusion gathered from comparing different methods and limitations of methodologies derived from Artificial Intelligent (AI) tools.
# Table of Content

Section 1 ..............................................................................................................................................................1  
1. Introduction: ......................................................................................................................................................2  
2. Literature Review: ............................................................................................................................................4  
3. Methodology:.....................................................................................................................................................5  
   3.1. Variation Analysis for Barnett Shale: ...........................................................................................................5  
   3.2. Variation Analysis for Eagle Ford Shale: ......................................................................................................8  
   3.3. Application of Fuzzy Logic for correlating PVT properties in Barnett Shale: ........................................11  
   3.4. Application of Fuzzy Logic for correlating PVT properties in Eagle Ford Shale: ........................................14  
4. Results/Outcomes: .........................................................................................................................................17  

Section 2 ...........................................................................................................................................................18  
1. Introduction: ....................................................................................................................................................19  
2. Literature Review: .........................................................................................................................................23  
3. Methodology:....................................................................................................................................................24  
   3.1. Gamma Ray Log: .......................................................................................................................................26  
   3.2. Resistivity Log ..........................................................................................................................................27  
   2.3. Density Log ...............................................................................................................................................29  
   2.4. Neutron Log .............................................................................................................................................29  
   2.5. Caliper Log ...............................................................................................................................................31  
   3.6. Clustering Analysis ................................................................................................................................32  
3. Results/Outcomes: .........................................................................................................................................45  

Section 3 ...........................................................................................................................................................46  
1. Introduction: ....................................................................................................................................................47  
   1.1. Regression Model for ROP Prediction: .......................................................................................................48  
   1.2. Smart Proxy for SAGD Simulation: ............................................................................................................48  
2. Literature Review: ..........................................................................................................................................49  
3. Methodology: ...................................................................................................................................................50  
   3.1. Predictive Model for ROP: ..........................................................................................................................50  
   3.2. Smart Proxy for SAGD Reservoir Model: ...................................................................................................69  
4. Results/Outcomes: ........................................................................................................................................75  

References:........................................................................................................................................................76
List of Tables

Table 1 Well Log Provided.................................................................................................................. 24
Table 2 Common Caliper Log Responses and their possible causes.............................................. 31
Table 3 Typical ranges of log properties for volumetric compositional analysis.......................... 37
Table 4 Model Architecture .............................................................................................................. 41
Table 5 Model Architecture .............................................................................................................. 55
Table 6 Details of Smart Proxy ......................................................................................................... 70
List of Figures

Fig. 1 Artificial intelligence and Major Subdivisions ................................................................. ix
Fig. 2 Ternary composition diagram of Barnett Shale fluid ......................................................... 5
Fig. 3 TVD (ft) versus Initial OGR (bbl/MMScf) for Barnett Shale ........................................... 6
Fig. 4 API gravity versus Initial OGR (bbl/MMScf) for Barnett Shale ........................................ 6
Fig. 5 C7+ WS (mol%) versus Initial OGR (bbl/MMScf) for Barnett Shale .............................. 7
Fig. 6 Ternary composition diagram of Eagle Ford Shale fluid .................................................. 8
Fig. 7 TVD (ft) versus Initial OGR (bbl/MMScf) for Eagle Ford Shale ....................................... 9
Fig. 8 API gravity versus Initial OGR (bbl/MMScf) for Eagle Ford Shale ................................... 9
Fig. 9 C7+ WS (mol%) versus Initial OGR (bbl/MMScf) for Eagle Ford Shale ....................... 10
Fig. 10—Fuzzy sets for different producing zones of Barnett Shale on the basis of OGR (bbl/MMScf) ........................................................................................................................................ 12
Figure 11 Screen shot of Calculations for fuzzy logic sets ....................................................... 12
Fig. 12- Fuzzy Set Model for Barnett Shale; True Vertical Depth(ft), Stock-tank oil API gravity, Initial stock-tank gas-oil ratio (GOR) and C7+ composition in produced well stream versus Initial stock-tank oil-gas ratio (OGR) ........................................................................................................ 14
Fig. 13 Fuzzy sets for different producing zones of Eagle Ford Shale on the basis of OGR (bbl/MMScf) ........................................................................................................................................ 15
Fig. 14 Fuzzy Set Model for E Shale; True Vertical Depth(ft), Stock-tank oil API gravity, Initial stock-tank gas-oil ratio (GOR) and C7+ composition in produced well stream versus Initial stock-tank oil-gas ratio (OGR) ........................................................................................................... 16
Fig. 15 An example of Clustering through KMeans Algorithm ................................................... 20
Fig. 16 Network for Multi-Layer Perceptron ............................................................................. 21
Fig. 17 Cross Plot Diagram of all the well logs provided .............................................................. 25
Fig. 18 Gamma Ray Interpretation ............................................................................................... 26
Fig. 19 Resistivity/ Conductivity readings to corresponding formations .................................... 28
Fig. 20 Schematic nuclear-log responses for some common lithologies ...................................... 30
Fig. 21 Optimization of Cluster Numbers .................................................................................. 32
Fig. 22 Distribution of Clusters in Data ....................................................................................... 33
Fig. 23 Distribution of Clusters along the Depth of Well Log provided ..................................... 34
Fig. 24 Cluster Analysis of different Well Log Properties ........................................................... 36
Fig. 25 Equations used for Computational Analysis-Composition ............................................. 37
Fig. 26 Volumetric Distribution of Rock type and Porosity with respect to Depth (ft) ............. 38
Fig. 27 Training Well Log .......................................................................................................... 39
Fig. 28 Prediction Well Log ........................................................................................................ 40
Fig. 29 Distribution of labels in prediction data ......................................................................... 41
Fig. 30 Predicted Label distribution along depth ........................................................................ 42
Fig. 31 Cluster Analysis of different Well Log Properties .......................................................... 43
Fig. 32 Volumetric Distribution of Rock type and Porosity with respect to Depth (ft) ............. 44
Fig. 33 Correlation Matrix between variables of lab test data ................................................... 51
Fig. 34 Log Diagram for model parameters .............................................................................. 53
Fig. 35 ROP Prediction for Blind set ................................................................. 56
Fig. 36 Comparison of Predicted ROP with Real ROP ........................................... 56
Fig. 37 Difference between Predicted and Real ROP in case of Bit balling .............. 57
Fig. 38 Log of Drilling parameters in case of Bit Balling ...................................... 58
Fig. 39 Relationship Matrix between Drilling parameters in case of Bit Balling ........ 59
Fig. 40 Comparison of relationship between Different Pressure and ROP/Torque (Before and after Bit Balling) ................................................................. 60
Fig. 41 Relationship between Independent variables (RPM, WOB, GPM) with Dependent variables (ROP and Torque) before and after Bit balling ................................................................. 61
Fig. 42 Relationship between ROP and Torque before and after Bit Balling ........... 62
Fig. 43 Change in ROP with respect to time (Before and after Bit Balling) .............. 63
Fig. 44 Log of Drilling Variables for Field case .................................................. 65
Fig. 45 Relationship matrix between Drilling Variables (Field case) ....................... 66
Fig. 46 Prediction of ROP for Blind Set ............................................................... 67
Fig. 47 Comparison of Predicted and Real ROP ................................................... 67
Figure 48. The tier system with the 4 cell in surface contact with the focal cell .......... 71
Figure 49 Areal comparison between Real and Predicted Pressure along with Error (%) - 4 Time Steps (18th, 34th, 39th and 55th Time Steps) ................................................................. 72
Figure 50 Areal comparison between Real and Predicted Oil Saturation along with Error (%) - 4 Time Steps (18th, 34th, 39th and 55th Time Steps) ................................................................. 73
Fig. 51 Pressure Distribution at time step=120 Days ........................................... 74
## Nomenclature

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>TVD</td>
<td>True Vertical depth (ft)</td>
</tr>
<tr>
<td>PVT</td>
<td>Pressure, Volume, Temperature</td>
</tr>
<tr>
<td>LRS</td>
<td>liquid Rich Shale</td>
</tr>
<tr>
<td>API</td>
<td>American Petroleum Institute</td>
</tr>
<tr>
<td>OGR</td>
<td>Oil Gas Ratio (bbl/MMScf)</td>
</tr>
<tr>
<td>Bbl</td>
<td>Reservoir Barrel</td>
</tr>
<tr>
<td>MMScf</td>
<td>Million Surface cubic Feet (unit of gas volume)</td>
</tr>
<tr>
<td>GOR</td>
<td>Gas Oil Ratio (Scf/bbl)</td>
</tr>
<tr>
<td>MLP</td>
<td>Multi-layer Perceptron</td>
</tr>
<tr>
<td>GR</td>
<td>Gamma Ray</td>
</tr>
<tr>
<td>PE</td>
<td>Photoelectric Effect</td>
</tr>
<tr>
<td>ROP</td>
<td>Rate of penetration (ft/hr)</td>
</tr>
<tr>
<td>SAGD</td>
<td>Steam Assisted gas Drainage</td>
</tr>
<tr>
<td>2D</td>
<td>Two Dimensional</td>
</tr>
<tr>
<td>WOB</td>
<td>Weight on Bit (klb)</td>
</tr>
<tr>
<td>RPM</td>
<td>Rotation per Minute</td>
</tr>
<tr>
<td>GPM</td>
<td>Gallons per Minute</td>
</tr>
<tr>
<td>DOC</td>
<td>Depth of cutting (inches)</td>
</tr>
<tr>
<td>DiffPress</td>
<td>Differential Pressure (Psi)</td>
</tr>
<tr>
<td>SPP</td>
<td>Standpipe Pressure (Psi)</td>
</tr>
</tbody>
</table>
Acknowledgments

And that there is not for man except that [good] for which he strives (53:39).

This thesis marks not only as a key milestone, turning pages away from yet another great chapter of my life but also as the beginning of a journey to several stellar possibilities. My two years at West Virginia University have left deep imprints on my life for all the good reasons and have also endowed me with a knowledgeable experience that will certainly open new doors of professional growth on me. Foremost, I would like to extend my thanks to the Almighty for giving me the strength to carry on this project and for blessing me with many great people who have been my greatest support in both, personal and professional life.

My earnest thank goes to the US Fulbright Scholar Program, without whom none of this could ever be possible. The assistance provided by them during my program in terms of legal and financial issues made it possible for me to focus comprehensively on attainment of my degree. I would also like to thank Dr. Ali Takbiri Borujeni for providing me with an opportunity to work and guiding me throughout the course of this program. I would also like to mention Dr. Reza Rahmani and Dr. Hamid Rahnema for providing me timely data and evaluating my progress to improve the outcomes of this work.

When I joined this department in fall 2016, I was naive to the world of data sciences and analytics, so if it was not for Dr. Shahab D. Mohaghegh, I could have never discovered an entirely different universe of data driven technologies. I would also like to express my gratitude to Dr. Ebrahim Fathi, whose critical analysis gave this work a more practical aspect. I would also like to thank my colleagues at WVU, Daniel and Dennis for being very helpful and providing me with the basis to start this work, Nupur, Dalal and Muhammet for giving immense emotional support during this whole journey.

Last but not the least, I would like to express my love and regards to my parents who have inspired me to keep growing and never giving up on my dreams. If it was not for their support, trust and belief I could not have achieved anything in my life. I would also like to mention my siblings Noor and Tauseef, who look upon me as a source of inspiration, which is a driving force for me to work harder and striving to be better.
Preface

Artificial Intelligence is relatively a younger subfield of Computer Sciences. The field of study was emerged in pursuit of creating better machines that can depict the human intelligence as closer as possible. Unfortunately, Artificial Intelligence (AI) is broad field with vaguely defined borders and these borders can be controversial. At the same time when different subfields of AI has been used extensively in different sectors, less and less focus has been given to defining strict rules and boundaries for use of the technology. Since it is still evolving, many times a user comes across the situation where the use of the technology is open to many interpretations.

The most popular school of thought believes that AI is the mother field of major subfields such as Fuzzy Logic and Machine Learning (Other minor field includes Natural language Processing). The fuzzy logic has been used separately or in combination with neural networks to generate different algorithms. Whereas, Machine learning or Statistical machine learning is the tool that utilizes the concept of Data Science/Data Analytics. The conventional applications of machine learning however are simply based on statistics, that provides optimized mathematical solution, provided the large amount of data. The functionality of conventional machine learning algorithm leans more towards “learning” rather than “intelligent”.

The currently hot topic of AI, is deep learning, which itself comes under the category of machine learning. The way it differs from conventional machine learning algorithm is that it mimics to biological neurons. The two very famous deep learning networks are convolution neural networks (CNN) and long short term memory (LSTM).

In this work we have utilized the Fuzzy Logic, Unsupervised Machine Learning (KMeans Clustering) which is a part of pure statically machine learning process and Multi-layer Perceptron (Classification and Regression) which is a feedforward backpropagation artificial neural network (ANN), arguably comes under the area of Deep Learning.

![Fig. 1 Artificial intelligence and Major Subdivisions](image)
Section 1
Application of Fuzzy Logic

Geospatial Analysis of Compositional Variations across Eagle Ford and Barnett Shale Plays
1. Introduction:

The published study under the name of “Geospatial Analysis of Compositional Variations across the Eagle Ford and Barnett Shale Plays (SPE-181702-MS)” was provided to us for extended analysis.

The above-mentioned work is an investigation of Liquid Rich Shales in Barnett and Eagle Ford. Data such as TVD, Areal Co-ordinates, and PVT properties was used to categorize the production from Eagle Ford and Barnett Shale into different reservoir zones such as Dry/Wet gas, Condensate (rich/lean) gas reservoirs.

Due to considerable variation in PVT properties with respect to depth and geospatial area it was not easy to see distinct boundaries between these regions, i.e. Dry-wet, Intermediate and Liquid Rich using conventional methods. The Fuzzy set theory was applied to deal with the ambiguity of distinct boundaries and visualize effect of PVT properties on different fluid types in both Barnett and Eagle Ford Shale.

Liquid-rich Shale (LRS) reservoirs such as the Eagle Ford Shale and the northern Barnett Shale have high potential to produce liquid hydrocarbons. This requires approaches to maximize and maintain liquid rate. As a result, different methodologies have been devised to understand the production and production decline in Liquid Rich Shales.

Kumar et al., (2013) mentions that Shale being a source rock, has very low relative permeability, this inhibits migration and fluid properties across the formation have higher heterogeneities. As a result, fluid composition not only varies within the reservoir but also with respect to time as pressure drops along the production.

Tian et al. (2013) also points out the variation in fluid properties on Eagle Ford Shale. Their studies focused on thermal maturity and depths, showing that Eagle ford has lithology and TOC that are highly varied with respect to depth and areal extents.

For better understanding of fluid behavior in LRS, fluid sampling is an essential procedure. However, sampling operations for LRS reservoirs are more difficult compared to conventional reservoirs. For example in conventional reservoir well stimulation such as acidization helps in stabilizing the flow rate and hence aids in collecting reservoir fluid samples that are supposedly better representation of reservoir condition. This is not the case for Shale reservoirs. Kumar et al., (2013) noted that sampling is one of the biggest challenges in LRS reservoirs. Since fluid samples are recommended to be taken in early life of a reservoir, whereas clean of hydraulic fracturing takes longer time, this gives rise to compositional changes in reservoir between sampling period and complete clean-up of near well bore. Both Whitson and Miranda further recommend that not only should samples be taken early but also with minimal drawdowns. Miranda et al., (2014) recognizes that low drawdowns may result in producing only fracturing fluid for an extended period of time, delaying any financial benefit from a
newly drilled well. Whitson et al., (2012) also appreciates potential operational challenges but still suggest this approach as it provides the best chance to obtain representative samples.

Since the samples collected are mostly not the true representation of reservoir’s in-situ fluid, there are possibilities of misrepresentation of fluid type while classification. The presence of two phase fluids in sample, even after bringing it in equilibrium with respect to reservoir’s temperature and pressure is an indication of non-representative fluid. The sample collected in such conditions has leaner components and therefore volatile oils may end up classifying as retrograde gases.

This works aims to investigate further about the compositional variation in Liquid Rich Shale reservoirs such as Barnett and Eagle Ford. Ternary diagrams were developed to find out better representation of different fluid types in the respective Shale formations. The work also shows variation in fluid properties such as C7+, API and initial Oil Gas ratio. We used the same data produced in earlier stages of this work to build a property window for reservoir fluid types, presumably present in respective reservoirs, using Fuzzy Logic Theory.
2. Literature Review:

Fuzzy Logic is application of recognized statistical techniques. When conventional techniques deal with absolute values, the fuzzy logic provides room for carrying inherent error term through the calculation rather than eliminating or minimizing it.

It has been extensively used in different engineering and soft-computer science communities, primarily, because it adds up a ductile connection between input and outputs by making the decision rely more on degrees of truth of the analysis instead of acting like on-off switches. The mathematicians have contradicted view about Fuzzy Logic because of it non-bivalence nature. This is exactly the reason why Fuzzy Logic is very useful in problem where mathematical bi-valence has limitations.

In Petroleum Engineering, Fuzzy Logic has been widely used for last two decades. The application is used for well log analysis, petrophysics, reservoir simulation, production analysis etc. Cuddy (1997) have successfully applied the methodology to petrophysics problem and compared the results with conventional methods. The November 2000 issue of SPE distinguished authors series have extensively discussed the application of Fuzzy Logic for candidate well selection by Mohaghegh (2000). Sung. H (2000) has also showcased the use of Fuzzy logic in areas of input importance ranking and log curve interpretation. In Anifowose (2013) work, he has utilized the combination of Fuzzy logic and Machine Learning to develop a hybrid model for prediction of reservoir properties. These works provide detailed methodology of application of Fuzzy Logic in different areas of petroleum engineering and give a brief comparison of using conventional methods for solving the same problems.
3. Methodology:

The data used for the study consisted of 97 wells in Barnett Shale and 153 wells in eagle Ford. The PVT reports used in the earlier part of this work includes composition and density of separator gas, liquid, well stream and advanced analysis such as Constant Composition Expansion (CCE) and Constant Volume Depletion (CVD) reports. The additional data includes but not limited as:

1. Producing formation
2. Latitude and longitude of wells
3. Well stream, separator gas and separator liquid composition with special attention to C7+ compositions
4. API gravity of the condensate
5. True Vertical Depth
6. Initial gas and oil production volumes (first full month) to calculate OGR

3.1. Variation Analysis for Barnett Shale:

The PVT reports provided from Barnett Shale shows two phase fluids, and hence the sampled fluid is not a good representation of in-situ fluid. To overcome this issue, multi-contact recombination was performed to have better representation of in-situ fluid. The results produced from the process helped in generating the below Ternary Diagram (Figure 1) that provide the basis of classifying reservoir fluids in Barnett, into three major classes, dry/wet gas, intermediate, liquid rich.

Fig. 2 Ternary composition diagram of Barnett Shale fluid
The TVD, corrected C7+, API gravity were plotted against initial OGR for the Barnett field. These are illustrated in Fig. 2, 3, 4 shows a correlation of C7+ vs. initial OGR. Overall, a little more than 100 of the 8000+ Barnett horizontal wells have been compared in this analysis and this is in no way reflects any chosen landing zone. A significant portion of data in a select area where different landing zone have been stimulated would be needed for such a study. The data for initial oil-gas ratios and well stream C7+ compositions are a scatter with depth. One would expect a similar trend for C7+ and initial OGR as each parameter can be a proxy for produced fluid density.

![Fig. 3 TVD (ft) versus Initial OGR (bbl/MMscf) for Barnett Shale](image1)

![Fig. 4 API gravity versus Initial OGR (bbl/MMScf) for Barnett Shale](image2)
Fig. 5 C7+ WS (mol %) versus Initial OGR (bbl/MMscf) for Barnett Shale
3.2 Variation Analysis for Eagle Ford Shale:

Eagle Ford is one of the unconventional play that covers the full spectrum of hydrocarbon thermal maturity windows from black oil to dry gas. Of the 153 PVT reports reviewed differences in static reservoir pressure and saturation pressures range from 1245 psi to in excess of 10,900 psi. This large degree of under-saturation increases the likelihood of obtaining representative samples but, in itself does not assure that collected fluids are indicative of in-situ fluids.

The validity of collected sample can be negatively impacted by several other considerations: from incorrect recombination GOR factors, sample contamination, leaks during transportation from field to laboratory etcetera. Typically duplicate samples are acquired and data validation steps are performed to increase confidence of representativeness. The vetting of collected fluid ought to be performed ahead of any further laboratory test. Any discrepancies during data validation are captured alongside additional laboratory tests.

A review of 153 Eagle Ford PVT Reports showed no indications of unrepresentativeness. The well stream heptane-plus percentage range from 2% to 14.4%. Based on accepted ranges for C7+, samples of both gas condensate and volatile oil samples were reviewed with 17 of the 153 being classified as volatile oil (C7+ > 12.5%). This is illustrated in Fig.5 below.

![Fig. 6 Ternary composition diagram of Eagle Ford Shale fluid](image)

TVD, C7+ and API gravity were tabulated against first full month OGR for Eagle Ford Wells. These are captured in Fig.6, 7, 8 below. Depth and Initial OGR can be correlated but there is a discontinuity.
As with Barnett data, Eagle Ford wells also show no correlation between initial OGR and depth or API gravity (Fig. 6 & 7). Fig. 8 shows that good correlation exists between Initial OGR and well stream C7+ despite Initial OGR not forming a strong correlation with depth. Further analysis of reservoir pressure shows the pressure gradient value is well above the hydrostatic pressure gradient and serves as confirmation that the Eagle Ford formation is over pressured. The source of this overpressure, whether due to hydrocarbon generation, rapid compaction or other sources is maintained across the entire portion of the evaluated field.

![Fig. 7 TVD (ft) versus Initial OGR (bbl/MMscf) for Eagle Ford Shale](image1)

![Fig. 8 API gravity versus Initial OGR (bbl/MMScf) for Eagle Ford Shale](image2)
Fig. 9 C7+ WS (mol%) versus Initial OGR (bbl/MMScf) for Eagle Ford Shale
3.3. Application of Fuzzy Logic for correlating PVT properties in Barnett Shale:

The conventional methodology used previously has its limitation when it comes to classification of different LGRs regions on the basis of PVT properties available for Barnett Shale. The generated results suggests that no effective or realistic relationship exists between depth and different PVT properties. However to avoid missing opportunity from high potential data, Fuzzy Logic Theory was utilized.

Fuzzy logic is comparatively a modern methodology, introduced by Lotfi A. Zadeh (1965). Since its origination, it has been used to generate rough solutions from the problems that by nature are vague or result of an imprecise process, where no crisp boundary exists between the classifiers. Contrary to what the name suggests, the methodology simplifies the relation between variables by defining the sets where each element of the set has certain degree of membership. Mathematically;

\[ \mu_{(A)}X \in [0, 1], \text{if } X \in A \]

Where, \( \mu_{(A)} \) is degree of membership?

X is an element in the system.

A is any set present in the system.

This implies that element X may have three possible memberships with respect to set A, either it can be 0, 1 or anything in between. The latter option provides the element an opportunity to have partial membership from more than one sets. Thus eliminating the need of pre-defined sharp boundaries, that may cause unrealistic conclusions, especially in case of erroneous data.

For understated problem, we approached same methodology by opting for three sets i.e. defined on the basis of OGR (bbl/MMscf). The sets were defined with following boundary conditions (Fig. 6)

1. Dry/Wet Gas: 0-15 bbl/MMScf
2. Intermediate: 15-120 bbl/MMScf
3. Liquid Rich: 120+ bbl/MMScf

We chose the ranges of fluid type based on the analysis provide by McMillan (2016). Since the sharing of boundaries between different sets is arbitrary, various combinations of shared boundaries were produced until we came across the fuzzy logic sets that produce more logical results and supports with previous analysis as closer as possible. For example since the existed knowledge about PVT proeprties supports that heavier the hydrocarbor lower will be the API gravity and higher will be the C7+ composition, therefore we iterated the set with different combination until such difference was visible in our genrted histograms.
Fig. 10—Fuzzy sets for different producing zones of Barnett Shale on the basis of OGR (bbl/MMscf)

The input of this fuzzy set model is all the PVT attributes used previously for the analysis, while output was defined as the type of reservoir fluid. For the wells having initial oil-gas ratio (bbl/MMscf) that lies between the shared boundaries of defined regions, membership function of the well was determined for the respective regions. Later, the same membership functions were used to split all the input parameters between these three fuzzy sets.

For example if a sample OGR is 110 bbl/MMScf, it has 0.27 membership in set “Liquid Rich” while 0.73 membership in set “Intermediate”. The same membership function was defined for it for other variables such as depth, API gravity, C7+ composition etc. If the same sample has TVD of 7000 ft, we would include TVD 7000x0.27=1890 ft for TVD in “Liquid Rich” and 7000x0.73=5110 ft for TVD in “Intermediate”.

Figure 11 Screen shot of Calculations for fuzzy logic sets
The results based on the Fuzzy set model infer that unlike conventional reservoirs, the fluid changes from dry/wet to intermediate however there is not an appreciable difference TVD from intermediate to liquid rich, as the depth, that means around same depth we can’t find either of this fluid and TVD is not the major factor to differentiate between these two fluid types. (Fig. 7a). The API behavior of reservoir fluids also implies that API decreases from dry/wet gas to intermediate and from intermediate to liquid rich (Fig 7b). The initial gas oil ratio in Barnett Shale suggests that liquid rich fluid has lowest amount of dissolved gas, in comparison to intermediate and dry/wet gas (Fig 7c). The C7+ components in the reservoir fluid are lowest in case of dry/wet gas, while rich liquids have most C7+ components (Fig 7d). This inverse relationship between initial gas oil ratio and C7+ illustrates a good correlation between the two parameters. At separator conditions, the amount of gas produced from different fluid types is similar to their initial gas oil ratio contrasts (Fig 7e). Though for liquid from C7+, there is only a minute difference between intermediate and liquid rich fluid (Fig 7f). This can be credited to liquid drop out in well bore, which is a common problem for liquid rich Shales (LRS).
To evaluate variation in different PVT properties in Eagle Ford’s categorized fluids i.e. Condensate and Volatile Oil, Fuzzy logic model was built with following boundary conditions for initial oil gas ratio (Fig. 12)

1. Condensate: 50-400 bbl/MMScf
2. Volatile Oil: 450-800 bbl/MMScf
The initial oil gas ratio limits for Eagle Ford’s Shale was applied as cited by Tian and Ayers (2013). The results generated through the methodology not only supports the previous analysis for vertical and areal variation, but successfully builds a meaningful relationship between depth and fluid types (Fig. 13). As previously seen for the case of Barnett Shale, the heavier reservoir fluids are present at deeper depths, while gas condensate can be found at shallower depths (Fig 13a). The API gravity of reservoir fluid increases from volatile oil to gas condensate (Fig 13b). The inverse relationship between initial gas oil ratio and C7+ of both the fluids is illustrated in Fig 13c and Fig 13d, where gas condensate has high initial gas oil ratio and lower C7+ components while volatile oil is rich with C7+ and has relatively lower initial gas oil ratios. On the surface gas condensate produces more gas from C7+ than liquid, while the volatile oil produces slightly more liquid from C7+ than gases (Fig 13e and 13f)
Fig. 14 Fuzzy Set Model for E Shale: True Vertical Depth(ft), Stock-tank oil API gravity, Initial stock-tank gas-oil ratio (GOR) and C7+ composition in produced well stream versus Initial stock-tank oil-gas ratio (OGR)
4. Results/Outcomes:

The results generated after the application of fuzzy sets, helped build a better relationship between the different reservoir parameters such as API, GOR and C7+ for understudy Barnett and Eagle Ford Shale. The analysis also provided a rough estimate of reservoir fluid properties for different reservoir fluid types in both reservoirs. This understanding of variation in PVT properties with respect to fluid type, aids to develop better reservoir models and build better strategies to maximize the liquid production in Liquid rich shales. However, the true representation of in-situ fluid in fluid samples is a pre-requisite of performing this analysis, or of any other analysis of such nature.

1. The use of fuzzy logic in this work is limited by expert’s experience. The relationship generated using the method, do not depend upon a physics-based equation, therefore it is crucial to develop a membership function for fuzzy logics, in consultation with subject expert. Increasing the granularity of the fuzzy set can provide better clustering and classification, hence resulting in more accurate analysis. For better results, it’s a thumb rule to use sufficient data.

2. It’s a recommendable practice to employ the application of data-based technologies such as fuzzy logic, pattern recognition, machine learning to solve for the inherent anomaly in Shale related problems.
Section 2

Application of Semi Supervised Learning
(Clustering and Classification)

Well Log Analysis
1. Introduction:

From Marcellus Shale Energy and Environment Laboratory, well logs were obtained and our objective was to find major lithological zones in the area with respect to depth. The technique used for the purpose is a combination of both supervised and unsupervised learning. The first part of the work is to identify the different targets in well logs using unsupervised learning. Once the clusters are formed, the results were confirmed with existing knowledge of well log analysis. The second part of the work uses the identified zones as a training set for supervised learning and predict similar zones in further well logs.

To understand Unsupervised Learning, it’s very essential to understand how it differs from supervised learning. Supervised Learning is the process where we have input variables and output variables and we use the algorithm to develop mapping function from input to find out the output, such as:

\[ y = f(X) \]  \hspace{1cm} (1)

The objective is to optimize this mapping function to predict the new output variable, given the new input data. The name of the learning process is self-explanatory, it learns from the already existing examples. The optimization generally stops when the mapping function achieves acceptable accuracy.

Majority of Machine Learning algorithms used for practical purposes classify under the category of supervised learning. In contrast to that, in Unsupervised Learning we only have input data and no corresponding output variables. It’s called Unsupervised because the algorithm learns on its own and comes up with interesting structure and interpretation of given data. The two most broadly classified problems under the Unsupervised Learning are Association and Clustering.

Clustering is the grouping of a particular set of objects in a manner based on their similar characteristics so that they aggregate in a group with other objects of similar nature, relative to that of other groups. These groups (clusters) are identified by distinguishable features that are common among the member objects of that cluster. It is one of the key techniques used in exploratory data mining which implements a specific join algorithm, based on required information analysis, to partition a given set of data. This methodology is particularly important in processing large data set of information to produce many distinct types of data. It has applications in many fields including image processing, machine learning, climatology, robotics, pattern recognition and data compression etc.

The general guideline for using the available algorithm in Scikit-learn libraries, provides a framework where one can choose the algorithm depending upon the structure of data and desired outputs. For Our problem we chose to use K-Means and DBSCAN method. The purpose of choosing two methods is to compare the results.

K-means clustering is a type of unsupervised learning algorithms which is used to classify unlabeled data into a fixed number of clusters (i.e. k). The key objective of using this methodology is to minimize the intra-cluster volatility i.e. the squared error function. The algorithm then works iteratively to assign each data point to one of the k clusters based on their feature similarity. This process involves four
steps, with the first one being to randomly select $k$ points as cluster centers. Theron, it assigns the data points to their closest cluster center by way of the Euclidean distance function and then calculating the centroid (i.e. mean) of all objects in each cluster. The difference between the previous and the recalculated centroids is computed and the steps are repeated over and over again until this value falls short than a threshold. K-means is equivalent to the expectation-maximization algorithm with a small, all-equal, diagonal covariance matrix.

![K-means clustering on the digits dataset (PCA-reduced data)](image)

**Fig. 15** An example of Clustering through KMeans Algorithm

*Classification* is the process of assigning new observations to a defined set of categories in accordance to derived characteristics from a training set. A classification model attempts to forecast the value of one or several outcomes based on the input data. The uses of classification can be witnessed in daily lives for e.g. when filtering emails as being spam or not spam, when categorizing some banking transaction data as being fraudulent or genuine or when assigning a diagnosis to patient based on his blood tests. This is one of the applications of pattern recognition and is used an important tool in machine learning. In the world of machine learning, classification is regarded as a subcategory of supervised learning where a predetermined training set is used against the categorization of observations.

The Scikit-learn libraries offer many algorithms for classification of data. Few of the popular classification algorithms are SVM, SGD, Gaussians, and MLPs. For this work, we have chosen Multi-Layer Perceptron method to classify our data.
Multi-layer Perceptron (MLP) is a supervised learning algorithm that learns a function by training on a dataset, where $m$ is the number of dimensions for input and $o$ is the number of dimensions for output. Given a set of features $X = x_1, x_2, \ldots, x_m$ and a target $y$, it can learn a non-linear function approximator for either classification or regression. It is different from logistic regression, in that between the input and the output layer, there can be one or more non-linear layers, called hidden layers. Figure 1 shows a one hidden layer MLP with scalar output.

The leftmost layer, known as the input layer, consists of a set of neurons $\{x_i | x_1, x_2, \ldots, x_m\}$ representing the input features. Each neuron in the hidden layer transforms the values from the previous layer with a weighted linear summation $w_1x_1 + w_2x_2 + \ldots + w_mx_m$, followed by a non-linear activation function $g(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ like the hyperbolic tan function. The output layer receives the values from the last hidden layer and transforms them into output values.

Class MLPClassifier implements a multi-layer perceptron (MLP) algorithm that trains using Backpropagation.

MLP trains on two arrays: array $X$ of size $(n\_samples, n\_features)$, which holds the training samples represented as floating point feature vectors; and array $y$ of size $(n\_samples,)$, which holds the target values (class labels) for the training samples:

After fitting (training), the model can predict labels for new samples:
MLP can fit a non-linear model to the training data. clf.coefs_ contains the weight matrices that constitute the model parameters:

MLP trains using Backpropagation. More precisely, it trains using some form of gradient descent and the gradients are calculated using Backpropagation. For classification, it minimizes the Cross-Entropy loss function, giving a vector of probability estimates $P(y|x)$ per sample $x$:

Currently, MLPClassifier supports only the Cross-Entropy loss function, which allows probability estimates by running the predict_proba method.

MLPClassifier supports multi-class classification by applying Softmax as the output function.

Further, the model supports multi-label classification in which a sample can belong to more than one class. For each class, the raw output passes through the logistic function. Values larger or equal to 0.5 are rounded to 1, otherwise to 0. For a predicted output of a sample, the indices where the value is 1 represents the assigned classes of that sample:

There are some problems where input data has some labelled outputs while the rest of data is open to interpretation. These problems are solved by Semi Supervised Learning algorithm. The use of Supervised and Unsupervised Learning, in earlier or later part of such problems depends upon the nature of problem and user’s discretion. The classic example of such algorithm are photo archives in our cell phones.
2. Literature Review:

The traditional method used in the classification of facies from well data has been the assignment of litho-facies by human interpreters who manually input data hence consuming a lot of time. In order to resolve this issue of tediousness and minimizing error, several alternative approaches have been designed by various researchers across the globe. One of the first such approaches has been the use of classical multivariate statistical models to classify facies (Wolf et al., 1982; Busch et al., 1987).

Thereon, Wolf et al. (1982); Busch et al. (1987) discovered a methodology to apply their knowledge of neural networks for the classification of rocks (Baldwin et al., 1990; Rogers et al., 1992). However, there has been a mushroom growth in the field of big data and also an exponential increase in the computational speed, which led to a reinvigoration in the use of sophisticated machine leaning technologies for this purpose. This technology has been of an increasing use in various fields ranging from the geological community to the mathematical as well as the astronomical community. Similarly petroleum researchers have taken a keen interest in exploring this field to use them in solving their problems in an effective manner (Smith and Treitel, 2010; Zhang et al., 2014; Zhao et al., 2015; Kobrunov and Priezzhev, 2016).

In relation to this problem, Hall (2016) proposed a Geophysical Tutorial whereby he applied common techniques of machine learning for the classification of facies. He used a dataset of seven wireline logs and used them to interpret and associate facies extracted from the wells of Hugoton Gas Field in Kansas (Dubois et al., 2007), to forecast geologic facies in two additional wells in the field based on their wireline measurements.

In the pursuit of this, he also organized a competition by inviting readers to compete in exploring machine learning for finding an alternative solution for the same task. Results of the machine learning contest are resumed in a paper written by the organizers (Hall and Hall, 2017) and all submitted code is available online.
3. Methodology:

For well log analysis two wells were selected. In these two wells, following eight log properties were shortlisted to use in process:

1. Gamma Ray Log (API)
2. Temperature Log (Deg F)
3. Caliper Calibration Log (in)
4. Neutron porosity Log (CF/CF)
5. Density Porosity Log (CF/CF)
6. Density Log (g/cm3)
7. Resistivity Log (Ohm.m)
8. Photoelectric Log

The first well was used for unsupervised learning and for the second well we applied supervised learning by using clustered data as training and predicting clusters for new dataset. The detailed information of each well is as follows:

<table>
<thead>
<tr>
<th>Well Name</th>
<th>Steps</th>
<th>Start(ft)</th>
<th>Stop(ft)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well No. 1 (Training Well)</td>
<td>0.5</td>
<td>70.5</td>
<td>8045</td>
</tr>
<tr>
<td>Prediction Well</td>
<td>0.5</td>
<td>63.5</td>
<td>7866</td>
</tr>
</tbody>
</table>

Table 1 Well Log Provided

The properties stated above were reported in same units for all both wells. Here the steps were the depth difference between two readings of logs.

For our understanding, we developed a pair plot between these eight properties. Apart from density of formation and Density porosity, which is the function of formation density, there was no visible relationship between the selected parameters (Fig. 17).
Fig. 17 Cross Plot Diagram of all the well logs provided
3.1. Gamma Ray Log:

The gamma ray log measures natural gamma ray emitting from a formation due to the presence of potassium-40 and the isotopes of the Uranium-Radium and Thorium series.

Working Principle:

The gamma ray detector is in the form of a scintillation counter which is composed of a thallium activated single sodium iodide. When a gamma ray strikes the sodium iodide a small flash is produced which is amplified by a photomultiplier made of a photocathode and a series of anodes. The flash now hits the photocathode number of primary electrons are produced. For every electron that hits the anode, a number of secondary electrons are emitted which are accelerated towards the next anode, where each of the secondary electrons produce even more secondary electrons. A current is generated that is amplified by conventional amplifiers.

The flash of light and the number of primary electrons is proportional to the energy of the gamma ray, the final current from the scintillation counter is also proportional to the energy of the incident gamma ray.

Gamma Ray log is useful for discrimination of different lithology when combined with other logs.

Interpretation:

Depending upon the responses from Gamma Ray log following interpretations are made:

![Fig. 18 Gamma Ray Interpretation](image-url)
3.2. Resistivity Log

Resistivity log records resistivities (or conductivities) of the formations, mostly along spontaneous potentials (SPs). These resistivities are product of formation conductivities and formation water salinity.

Working Principle:

The main objective of resistivity log is to determine Rt (resistivity of formation) and Rxo (resistivity of invaded zone) which is achieved by combining below three depths of investigation.

The deep-reading focused devices include:

- The deep laterolog device (LLd)
- The deep induction (ILd) device

Medium-depth devices include:

- The shallow laterolog (LLs)
- The medium induction (ILm)

The shallow devices include:

- The microresistivity device
- The spherically focused log (SFL)

Interpretation:

Depending upon the responses and information collected following interpretations are made:
As a thumb rule, resistivity of different formation and fluids can be descending written as:

Resistivity of Saline water

Resistivity of Formation bearing water

Resistivity of formation bearing water and hydrocarbons
2.3. Density Log

As the name suggests, density log measures the bulk density of the formation. Its measures porosity of the formation as a function of density. It also detects of gas-bearing formations and evaporites.

*Working Principle:*

A density log works like other active open hole tools by sending gamma rays into a formation and detecting scattered back. The sondes use a Cesium-137 source that emits gamma rays of a high energy. The scattered back gamma rays that detectors receive, are function of electron density of the formation. The lower electron density implies lower bulk density.

When used in combination with neutron log, density log is a very god method for lithological identification.

2.4. Neutron Log

The neutron log detects amount of hydrogen atoms in a formation thus is mainly used for determining formation porosity.

*Working Principle:*

The neutron logging tool emits neutrons that travels into the formation, loses energy and resulting in high energy gamma rays. The scattered neutrons lose energy due to reaction with hydrogen atoms. Thus, more the amount of hydrogen atoms lower will be the energy of receiving neutrons. It means that in porous formation, count on receiving end will be low or vice versa.

The cross plot between Neutron Log and Density log helps to identify evaporates, distinctive lithologies, shaly formations, clean formations, etc.

*Interpretation:*

Neutron log is viewed along density and PE log for better interpretation.
Fig. 20 Schematic nuclear-log responses for some common lithologies
2.5. Caliper Log

The caliper log is a passive measurement technique for open hole logs. It’s used for measuring the diameter and shape of a borehole.

Working Principle:

In the two arms tools, the borehole diameter is measured. Due to tectonic activities, the well bore can be oval, and the bore hole diameter can be more or less that the bit diameter. In case the well bore is oval, arms of caliper lock into the long axis of the well bore cross-section that gives larger value of well bore diameter. For such cases we need more than two arms calipers.

For dual arm caliper, the two opposites bore hole diameter is calculated in two axes perpendicular to each other as in the case of the Borehole Geometry Tool (BGT). Such tools can open to 30 inches and give two readings that are independent to each other. Along with that the tool allows volume measurement and has special sensors to from azimuth and dips. In more advanced cases 30 arms caliper also present that enables the operators to plot the detail well bore trajectory.

Caliper log provides lithological information.

Interpretation:

Depending upon the responses and information collected following interpretations are made:

<table>
<thead>
<tr>
<th>Hole Diameter</th>
<th>Cause</th>
<th>Possible Lithologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>On Gauge</td>
<td>Well consolidated formations Non-permeable formations.</td>
<td>Massive sandstones</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Calcereous shales</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Igneous rocks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metamorphic rocks</td>
</tr>
<tr>
<td>Larger than Bit Size</td>
<td>1. Formation soluble in drilling mud.</td>
<td>1. Salt formations drilled with fresh water.</td>
</tr>
<tr>
<td></td>
<td>2. Formations weak and cave in.</td>
<td>2. Unconsolidated sands, gravels, brittle shales.</td>
</tr>
<tr>
<td>Smaller than Bit Size</td>
<td>1. Formations swell and flow into borehole.</td>
<td>1. Swelling shales.</td>
</tr>
</tbody>
</table>

Table 2 Common Caliper Log Responses and their possible causes
3.6. Clustering Analysis

The foremost step of starting unsupervised learning is to choose methods for clustering. For our work we chose K-Means and DBSCAN Method.

KMeans: To find the optimized No. of Clusters for KMeans clusters, we iterated for multiple no. of clusters and found inertia (which is inversely proportional to accuracy of model). By plotting no. of clusters and Inertia, we narrowed down our number of clusters to three after which there was no significant change in inertia, with increase in number of clusters (label 0, 1 and 2).

The selected model was used to predict labels for entire dataset. To understand the distribution of the labels in dataset, a histogram was generated. It’s safe to say that the most of data was classified in major two clusters, while only few data points were part of third cluster label. Then we color coded these labels and tagged them alphabetically to see the distribution of labels along the depth log.
The deeper formation shows high resistivity and gamma ray values, and characterized by the clustering as Cluster A, while the shallower formation low gamma ray value and resistivity while the photoelectric effects has higher value of 10. Between these two very noticeable contrasting geological zones we have a layer that whose properties are similar to Cluster B properties and but has different range of density, therefore the neutron porosity and density porosity also differ, very significantly. As soon as the log reads this zone the caliper log deflects to the higher value of 9 inches from initial value of 5 inches. The trend between neutron porosity and density porosity suggests that this zone can be hydrocarbon bearing zone and have significant porosity (Fig 22). The high density in Cluster C can be due to presence of either calcite or dolomite containing water (due to high neutron porosity) and high caliper log reading.
Fig. 23 Distribution of Clusters along the Depth of Well Log provided
The cluster labels A, B and C, signify the presence of major three lithological zones in the given well log. In Fig. 23, the ranges in which the values of different log varies within cluster can be analyzed. For Cluster A, the average density porosity is less than 0.1 while the average neutron porosity is around 0.1. The Cluster B which is a shallower formation shows very high density porosity due to lower formation density and therefore is not very reliable, while the neutron porosity is near 0. The Cluster C shows negative density porosity as a result of very high density while the neutron porosity ranges from 0.2-0.35.

Cluster A has very high gamma ray values of around 200 API, while Cluster B has lower value of around 50 API, which supports the theoretical concept of deeper formations having more radioactive elements as compared to shallower formations. Interestingly, Cluster C has covers a larger variety of gamma ray that majority varies from 50 API to 120 API. This could be due to the fact that these zones are not clean formations and have significantly thicker streaks of different formations i.e. of calcium carbonates and clays.

The average density of Cluster A is around 2.7 g/cm^3 while for Cluster B it is around 2.2 g/cm^3. The cluster C have significantly higher density of around 3.7 g/cm^3. The higher densities can be due to presence of heavy fluids (i.e. oil and water) along with heavier formations such as calcite (Limestone), dolomites etc.
Fig. 24 Cluster Analysis of different Well Log Properties
Once the clusters were analyzed based on the apparent properties, we performed the computational analysis for composition of well log based on the crossplots. The conventional method of finding out the matrix and porosity distribution was used to find out the detailed characteristics of clustered zones. The cross plots used in the computational analysis are:

1. Neutron Density Cross Plot
2. Density-Pe Cross Plot

![Fig. 25 Equations used for Computational Analysis-Composition](image)

The logs used for calculations are:
- Density Log
- Neutron Porosity Log
- Photoelectric potential
- Gamma Ray

The typical ranges for normalization was taken as:

<table>
<thead>
<tr>
<th></th>
<th>Quartz</th>
<th>Calcite</th>
<th>Dolomite</th>
<th>Porosity</th>
<th>Shale</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHIN</td>
<td>-1.5</td>
<td>0</td>
<td>2.5</td>
<td>100</td>
<td>30</td>
</tr>
<tr>
<td>RHOB</td>
<td>2.65</td>
<td>2.71</td>
<td>2.87</td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>Umma</td>
<td>4.79</td>
<td>13.77</td>
<td>9.05</td>
<td>1.95</td>
<td>7</td>
</tr>
<tr>
<td>GR</td>
<td>25</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>120</td>
</tr>
<tr>
<td>Unity</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3 Typical ranges of log properties for volumetric compositional analysis
According to the conventional method, Cluster A majorly constitutes of dirty shale, with streaks of calcite and very low porosity. While Cluster B is majorly a calcite with insignificant presence of shale formation. The Cluster C however is a zone of high porosity and is therefore can be a hydrocarbon/water bearing formation. The matrix of Cluster C constitutes of majorly calcite and streaks of shale.
Once the different zones were identified using clustering analysis, the whole log was trained to predict the similar zones in other well log. It can be seen from Fig. 25 that trained well log has similar lithological zones as well that was used for prediction (Fig. 26). The reason of this similarity is that both wells were drilled in vicinity and around same depth. However the thickness of identified lithological zones varies.

Fig. 27 Training Well Log
Fig. 28 Prediction Well Log
For the purpose of classification MLP Classifier was used. Before training the model, the data was scaled using MinMaxScaler Function. The MinMaxScaler algorithm performs relative scaling of the whole range of data with respect to its minimum and maximum value, mostly inform of zero to unity, or in some cases from –1 to 1. As a result, the maximum value of the range is usually scaled to the unit sizes. The use of scaling ensure that standard deviation is small and sparse data has no entries. The architecture of the model was as below:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value/Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutrons</td>
<td>50</td>
</tr>
<tr>
<td>solver</td>
<td>adam</td>
</tr>
<tr>
<td>Hidden layer size</td>
<td>50</td>
</tr>
<tr>
<td>Tolerance</td>
<td>0.000001</td>
</tr>
<tr>
<td>Maximum Iteration</td>
<td>1000</td>
</tr>
<tr>
<td>Learning rate</td>
<td>adaptive</td>
</tr>
<tr>
<td>Initial learning rate</td>
<td>0.01</td>
</tr>
<tr>
<td>Activation</td>
<td>relu</td>
</tr>
<tr>
<td>Momentum</td>
<td>0.9</td>
</tr>
<tr>
<td>Shuffle</td>
<td>True</td>
</tr>
</tbody>
</table>

Table 4 Model Architecture

The rest of MLP Classifier features were taken at default settings. The model was first fit to the training set and was iterated for hundred times. The mean of each run was taken as a predicted value.

Fig. 29 Distribution of labels in prediction data
The predicted labels against new well was later color coded and tagged alphabetically to see the distribution of labels along the depth log.

**Fig. 30 Predicted Label distribution along depth**

The model predicted the identified lithological zones within acceptable accuracy. The next step was to compare the properties of identified labels against that of trained cell. Different log properties were plotted with respect to their cluster labels to visualize the ranges of values and distribution of these properties (Fig. 29)
The overall porosity, let it be density or neutron porosity for each cluster corresponds to same value as in case of training set. While the gamma ray values differ slightly as compared to training wet. As compared to training set, all three clusters have lower gamma ray values in prediction set. This could be due to slight change in lithology or difference of resolution between logging tool used for the different wells. The photoelectric effect and resistivity ranges of all the three cluster remains same as they were in training set. The density of three different cluster labels remains same as they were initial in training except with a minute exception Cluster B which does not have any layer of density lower than 2 g/cm$^3$ anymore.
As done previously, we performed volumetric analysis of the prediction ell with respect to depth to find out the distribution different matrix types and porosity. The overall porosities of the three clusters differ as compared to the training set but the major matrix distribution in the clusters remain same, i.e. in Cluster A shale is the major formation along with calcites, in Cluster B, calcite is major formation with streaks of shale, Cluster C is similar to Cluster B with better porosity.

Fig. 32 Volumetric Distribution of Rock type and Porosity with respect to Depth (ft)
3. Results/Outcomes:

For many years before the invention of big data and learning geologists and geoscientists have been working with enormous data. But now when the large and complex geological data is readily available the machine learning tools have been widely used to explore and analyze this data. The availability of open source packages has made it easier to extract important information from geological data without investing months and using licensed software.

The results produced from this work provides confidence about using unsupervised learning to differentiate different lithological zones based on parameters of provided data. In the first part of this project we investigated a form of unsupervised classification called clustering where instead of teaching our network to recognize the geological zones identified by the geoscientists, we gave it the freedom to come up with its own classes/labels and then analyzed the similarities and differences with the human interpretation and results obtained from clustering.

In the later part of this work the identified zones were trained to see if the machine learning algorithm for classification problems can classify the blind data against the labels in which that data belongs. The results produced also confirm a capability of developed algorithm to generalize to new data.
Section 3

Application of Supervised Learning

Multi-Layer Perceptron Regression

Predictive Model for ROP and SAGD
1. Introduction:

Different problems in oil and gas industry are categorized as continuous value problems or in statistical terms “regression” problems. The regression analysis performed in such kind of problems is all about estimating the effect of each variable on another variable. These problems consists of a dependent or a set of dependent variables and one or more independent variable. For example, according to newton’s second law of motion, the acceleration of an object as produced by a net force is directly proportional to the magnitude of the net force, in the same direction as the net force, and inversely proportional to the mass of the object. In statistical model this law suggests that acceleration is a dependent variable, effected by independent variables such as Force and Mass. The relationship between these variables is build statistically through regression analysis. The physical equation developed mostly constitute of these dependent and independent variables, while in statistical model one more parameter unknown parameters are involved such that

\[ y \approx f(X, \beta) \]

Here, y is a dependent variable.

X is independent variable/s.

\( \beta \) is unknown variable.

Since the basis of engineering lies within physics, we have developed all the relationships between variables on different dependent and independent variables. While there have been huge contributions made in centuries to develop these relationships, there are still many problems where relationship between variables have not been understood completely or entirely. The statistical tools such as regression analysis enables to utilize the data provided to predict the dependent variable before even shortlisting the variable it depends upon. It also includes effect of variables that may/may not apparently effect the dependent variable. Since the statistical model gives more room for finding a solution with huge number of variables that affect the dependent variable, it has been used quite widely for problems with indefinite solutions.

The current favorite tool used by engineers and scientist to perform regression analysis is machine learning. The ability of machine learning tools to handle big data and performing deep learning helps perform the regression analysis, which itself is extensively data depend. The scikit-learn library we chose for regression is multi-layer perceptron regression.

In the last section we discussed multi-layer perceptron along with multi-layer perceptron classification in details. The difference between the regression and clustering using multi-layer perceptron is that in MLP Regression, the multilayer perceptron is trained using backpropagation, without activation function in layer(contrary to classification), this leads to use of identity function as activation function and the outputs are set of continuous values(contrary to labels in classification).
Background:

We chose two entirely different problems to perform regression analysis. The details of both problem are:

1.1. Regression Model for ROP Prediction:

The most common problem faced in drilling sticky-shale or loose sandstone is “balling up” of the drill bit. The bit baling can sometime increase the drilling duration considerably and in very rare cases can cause the unplanned pull out of hole. Hence, preventing this “bit balling” can result into better drilling operation. The direct indicator of bit balling is reduced ROP at increasing WOB. Therefore, ROP is the most important factor that determine the successfulness of any drilling project. The foremost objective of this work is to develop a predictive model for ROP, which would later on lead to optimizing it by studying effect of different drilling parameters.

For any Oil and gas project, drilling is the most expansive operation. The success of any drilling operation is based on three principles:

1. Increased Drilling Speed(ROP)
2. Lower overall cost
3. Maintaining Safety

But due to the uncertainty of formational geology and uncontrollable operation constraints, it is challenge for any drilling team to optimize drilling operation. It is essential to understand the effect of different drilling parameters on ROP, to predict and later optimizing the ROP. But due to inherent complexity of different mathematical and statistical model, the prediction of ROP has always been questionably inaccurate.

1.2. Smart Proxy for SAGD Simulation:

The reservoir simulation models developed to perform forecasting for production performance of reservoir under different producing condition are based on complex differential equations. The special industrial software are developed to run these models that simulate the provided conditions and forecast production/parameters. However the more complicated the reservoir and applied conditions are the longer it takes to simulate them. As the last part of this work, we tried to stretch the applicability of generated regression model and analyze how they perform in contrast to reservoir simulation using licensed software. The 2D reservoir model was generated and we simulated it to forecast for five years under the steam assisted gravity drainage stimulation.
2. Literature Review:

Previously, different methodologies have been devised that used direct or indirect approaches to evaluate ROP. ROP is a function of drilling parameters, drilling fluid type and most importantly the properties of rock being drilled. That’s why it’s a direct indicator of rocks mechanical property. This leads us to improve drilling parameters, bit design and fluid type to achieve desirable ROP in all kinds of formations.

Bourgoyne and Young (1974) came forward with the idea of evaluating ROP as a function of eight variables, where these parameters were the result of multiple regression of data. The equation developed by them was on valid for roller-cone bits. After their work many authors got inspired to connect ROP with different rock properties variables, because they were available using well logs. Even right now most companies make drilling designs especially choice of drill bits and drilling fluid depending upon rock mechanical properties that is acquired from well logs.

Howarth and Rowlands (1987) describes in their work how drillability of rock is a function rock mechanical property such as uniaxial compressive and tensile stress. Following that work, Kahraman (2002) utilized lab test data to correlate rate of penetration with brittleness of rock, which is also a function of uniaxial stress.

In reality, rock mechanical properties are not enough to predict ROP, because ROP is also the function of different real-time parameters such as drilling operations. To prove the concept many authors have presented their work to predict ROP as a function of drilling parameters such as weight on bit, RPM, Mud flow rate etc. But unfortunately even these drilling parameters are not enough to predict ROP, since there are many other uncertain parameters that affect the ROP and are essential in developing a predictive method for ROP. That’s why when the neural network started being popular, many people decided to go for this soft computing tool to predict ROP. The biggest advantage of using neural networks for such problem is that it offers the flexibility of including all the available information in developing a predictive method.

The relatively new work by Bahari and Seyed (2008) has utilized genetic algorithm for calculating different constant coefficients that can be used in conjunction with Bourgoyne and Young’s equation for developing an ROP model. Meanwhile many new researchers have been working on developing a ROP model based on artificial neural network. Our works is an initiative of same efforts, to assess limitation of artificial of neural network to develop ROP predictive model.
3. Methodology:

3.1. Predictive Model for ROP:

3.1.1. Model Inputs and Output:

The data provided for this work mostly consists of drill bit lab tests. The major variables available for creating the model are:

1. Charge Pressures (Psi)
2. Rotations per minute (RPM)
3. Mud Flow (gpm)
4. Weight on bit (klb)
5. Bore hole Pressure (Psi)
6. Swivel Pressure (Psi)
7. Choke Pressure (Psi)
8. Torque (klb.ft)
9. Penetration (in)
10. Depth of cutting (in)

Since the lab tests are the simulation of field operation, it’s our basic assumption that if we can model the lab data, we will be able to modify the same model for field data variables.

The heat map between the variables established that only pressures like swivel, borehole pressure and choke pressure are directly proportional to each other. Since torque is also the function of load, that’s why we decided to exclude torque from the model (Fig. 13).
Fig. 33 Correlation Matrix between variables of lab test data
Based on existing knowledge of drilling parameters and the heat map generated in earlier stages of work, we chose following parameters for training our predictive model:

1. Charge Pressure1 (psi)
2. Charge Pressure2 (psi)
3. Mud Flow (gpm)
4. Weight on Bit (klb)
5. Borehole Pressure (psi)
6. Swivel Pressure (psi)
7. Choke Pressure (psi)
8. Rotation per minute
9. ROP (ft/hr)

Before starting the algorithm, we removed half of the ROP manually to use the complete data file for training and then predicting removed ROP to assess the efficiency of the model. (Fig. 14)
Fig. 34 Log Diagram for model parameters
3.1.2. Pre-Processing of Data:

The Multi-layer perceptron regression algorithm was selected for developing the predictive model. The MLP models are very sensitive to scaling. To perform pre-processing of data we chose build in scikit-learn algorithm for MinMaxScaler. The MinMaxScaler algorithm performs relative scaling of the whole range of data with respect to its minimum and maximum value, mostly inform of zero to unity, or in some cases from –1 to 1. As a result, the maximum value of the range is usually scaled to the unit sizes. The use of scaling ensure that standard deviation is small and sparse data has no entries. Since the visualization of pair plot for parameters in Fig. 15, does not indicate any part of dataset as outlier. We did not apply any outlier detection on our dataset.
3.1.3. Training of Model:

For training, we used 50% of available data whereas remaining 50% was used as prediction set. The Multi-Layer Perceptron Regression model we developed had following architecture.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value/Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutrons</td>
<td>50</td>
</tr>
<tr>
<td>solver</td>
<td>adam</td>
</tr>
<tr>
<td>Hidden layer size</td>
<td>50</td>
</tr>
<tr>
<td>Tolerance</td>
<td>0.000001</td>
</tr>
<tr>
<td>Maximum Iteration</td>
<td>1000</td>
</tr>
<tr>
<td>Learning rate</td>
<td>adaptive</td>
</tr>
<tr>
<td>Initial learning rate</td>
<td>0.01</td>
</tr>
<tr>
<td>Activation</td>
<td>relu</td>
</tr>
<tr>
<td>Momentum</td>
<td>0.9</td>
</tr>
<tr>
<td>Shuffle</td>
<td>True</td>
</tr>
</tbody>
</table>

Table 5 Model Architecture

The rest of MLP Regressor features were taken at default settings. The model was first fit to the training set and was iterated for hundred times. The mean of each run was taken as a predicted value.
Once the values of ROP were predicted, we imported the removed values of ROP to compare our predicted ROP. The results for this case were promising.

![Fig. 35 ROP Prediction for Blind set](image)

We applied the same procedure for remaining lab tests data to see the applicability of our developed model. The results from different lab tests data are included in the glossary. There were few cases within lab data where our prediction pointed out some insights from the data. One of such cases is discussed below.

![Fig. 36 Comparison of Predicted ROP with Real ROP](image)
3.1.3.1. Cases with Bit Balling:

Bit balling is characterized as slowness of penetration rate. Many parameters contribute to slow rate of penetration, for example, formation characteristics, bit type, drilling fluid properties, drill bit hydraulics, operating conditions etc. Since not all of these parameters are covered in our training system, that’s why the model is unable to predict the bit balling using limited provided data. The prediction made in case of bit balling shows a difference between real and predicted value where predicted value increases with depth and weight on bit, while in actual ROP decreases gradually.

Fig. 37 Difference between Predicted and Real ROP in case of Bit balling

As any other case of bit balling it’s interesting to note that along ROP, Torque on bit also decreases contrary to its normal response with increasing depth. In normal response, as the drilling proceeds the ability of drill bit to drill the formation and the torque on bit increases.
Fig. 38 Log of Drilling parameters in case of Bit Balling
To see the general relationship between different variables in the current case, a heat map was generated that establishes that ROP and Torque are highly correlated. As expected, ROP and Torque share the same response with weight on bit.

**Fig. 39** Relationship Matrix between Drilling parameters in case of Bit Balling
Different parameters were plotted against each other to identify their relationship with ROP and Torque and to visualize how this relationship diverts in case of bit balling. Since the data provided to us is in milliseconds and there is enormous fluctuation in some of the variables. It is not advised to develop any relationship between variables. For the sake of visualization we compared Torque and ROP with different pressures first, i.e. Charge Pressure 1, Charge Pressure 2, Swivel Pressure, Choke Pressure, and Borehole Pressure (Fig XX). We do not see any noticeable difference in cases with bit balling and with bit balling except in the case of Charge Pressure 2. In the normal process, ROP and Torque are increasing with increments in Charge Pressure 2. While in the case of Bit Balling, the relationship becomes exactly opposite and the ROP keeps decreasing even while Charge Pressure 2 is incremented. For swivel pressure, ROP and Torque seems to share an inverse relationship with the variable but as bit balling starts ROP and Torque does not seem to be affected at all by any increment in swivel pressure.

![Fig. 40 Comparison of relationship between Different Pressure anad ROP/Torque (Before and after Bit Balling)](image)

However, when Torque and ROP were plotted against variables like weight on bit, RPM and Mud Flow, a better understanding of bit balling was achieved. The usual response of both ROP and Torque with RPM is inversely proportional while in case of bit balling both output variables turns independent of any changes in RPM. The general trend suggests that ROP and Torque subsequently increase with increase in Mud Flow (GPM), but as soon as bit balling happens there is no effect of Mud Flow on ROP and Torque any longer. In case Weight on bit, the ROP and Torque reacts to increase in weight on bit with almost same rate. The rate of change in ROP with respect to rate of change in WOB is almost equal. Though the same variables do not share similar relationship as bit balling starts. The ROP and Torque do change with WOB but the rate of change is considerably very low. Since the model is trained for responses in usual conditions, it cannot predict the bit balling.
Fig. 41 Relationship between Independent variables (RPM, WOB, GPM) with Dependent variables (ROP and Torque) before and after Bit balling

It is also interesting to see that the highly correlated ROP and Torque change their correlation as bit balling happens. Earlier the ROP and Torque are changing with same rate, while after bit balling the rate of change in ROP with change in Torque decreases.
Fig. 42 Relationship between ROP and Torque before and after Bit Balling
In the previous section of this work, we have applied clustering analysis on well log data and have separated out different geological sections from the well logs. The same methodology was applied to separate out zone of No Bit Balling and Bit Balling in given data set, where Bit Balling is very apparent.

![Fig. 43 Change in ROP with respect to time (Before and after Bit Balling)](image-url)

**Fig. 43 Change in ROP with respect to time (Before and after Bit Balling)**
3.1.3.2. Field Case Study:

Since we also had one case study from field and we wanted to extend the applicability of this regression model for field variables, we tested our model with field data. For field data our variables were:

1. Standpipe Pressure (Psi)
2. Torque (klb.ft)
3. Weight on Bit (klb)
4. Flow Rate (GPM)
5. Differential Pressure (Psi)
6. Rate of Penetration (ft/hr)
7. Total RPM

As done previously we removed half of the output data i.e. the ROP from the given file to test the predictive model for the rest of the data. The log plot for all variables provided the information of how each variable is changing on the course of drilling time. There is one noticeable period in the data where SPP, differential pressure, flow rate and Total RPM falls to zero for some time lesser than an hour, supposedly it’s the time when drilling stops and after they resume the drilling. This part of data lies in the prediction partition and we expect the regression model to predict what happens to ROP during that. All the variables were reported against elapsed time, the difference between each reading varies from 4-6 seconds.

Since the data for field is a bit different data that was available for lab data, it is recommended to see the comparative relationship between variables. The heat map generated for the purpose implies that Torque is highly correlated to RPM, differential pressure and weight on bit. While ROP is comparatively better correlated with RPM than any other variable in the data. This suggests that Total RPM affects Torque on bit and ROP in very considerable magnitude.

To have an idea about what nature of relationship each variable share with every other variable in our data, a pair plot was generated. The general trend in the plot suggests that variables other than SPP an differential Pressure, that have direct relationship, do not share any evident direct relationship.
Fig. 44 Log of Drilling Variables for Field case
Fig. 45 Relationship matrix between Drilling Variable (Field case)

Torque was not included as input for the data since it’s a resultant of drilling operation just like ROP. To train the model we chose exact model architecture and applied MinMaxScaler as a part of pre-processing. The model was iterated for 100 times and the mean of predicted result from each run was taken as a predicted value of ROP.
Fig. 46 Prediction of ROP for Blind Set

The results generated against blind part of the data were later compared with the actual values of ROP to assess the effectiveness of predictive model. The predictive ROP does not replicate exactly like actual ROP but is capable to capture the trend of ROP in prediction set. More interestingly the model has successfully predicted the ROP very accurately for the period where drilling was stopped.

Fig. 47 Comparison of Predicted and Real ROP
The results generated after applying the regression model on field case study strengthens the initially made assumption that if we can create a model to predict lab data, it can also be used for predicting real data from field. Further availability of field data will provide better confidence of capability of generated algorithm to be applied on general cases.
3.2. **Smart Proxy for SAGD Reservoir Model:**

3.2.1. **Data Variables:**

The data utilized for this work was generated using output file from a reservoir simulation software. The 2D Model had 50x30 cells, 41 properties and 60 time steps. The time steps were divided into months while the variables available in the files are:

1. Pressure (KPa)
2. Water Saturation
3. Oil Saturation
4. Gas Saturation
5. Temperature (Celsius)
6. Oil Relative Permeability
7. Gas Relative Permeability
8. Water Relative Permeability
9. Oil Fractional Flow
10. Gas Fractional Flow
11. Water Fractional Flow
12. Void Porosity
13. Fluid Porosity
14. Incremental Subsidence (m)
15. Depth to Top of Water Column (m)
16. CH4 Bubble Point Pressure (kPa)
17. Subsidence (m)
18. Vertical Displacement (m)
19. Implicit IMPES Map
20. Oil Viscosity (cp)
21. Gas Viscosity (cp)
22. Water Viscosity (cp)
23. Oil Molar Density (gmole/m3)
24. Gas Molar Density (gmole/m3)
25. Water Molar Density (gmole/m3)
26. Oil Mass Density (kg/m3)
27. Gas Mass Density (kg/m3)
28. Water Mass Density (kg/m3)
29. Fluid Pore Volume (m3)
30. Fluid Enthalpy (J/gmole)
31. Thermal Conductivity (J/m-day-C)
32. Heat Capacity (J/m3-C)
33. Heat Loss Rate (J/day)
34. Net Heater Accumulation (J)
35. Net Heater Rate (J/day)
36. Oil Resistance Factor
37. Gas Resistance Factor
38. Water Resistance Factor
39. Capillary Pressure (w/o) (kPa)
40. Capillary Pressure (o/g) (kPa)
41. Heavy Bubble Point Pressure (kPa)
3.2.2. Preprocessing of Data:

The data available in matrix was converted into columns where each column represented the property/parameter and the row represented the position of cell. Such files were developed for 60 different time steps.

Since each cell doesn’t work independent and is affected by the neighboring cells. Tier-1 neighbor cells were assigned against each cell and their properties were included so now each cell had 41 properties of its own and 4x41 properties of its neighbors.

3.2.3. Training of Model:

The model was trained with non-cascading approach. The model was trained for three different time steps (10\textsuperscript{th}, 40\textsuperscript{th} and 60\textsuperscript{th} Time Steps out of 60 Available Time Steps). The model was then used to predict pressure and saturation of oilf for timestep 18\textsuperscript{th}, 34\textsuperscript{th}, 39\textsuperscript{th} and 55\textsuperscript{th}.

<table>
<thead>
<tr>
<th>For Predicting Pressure</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Inputs</td>
<td>41-1=40</td>
</tr>
<tr>
<td>Number of Output</td>
<td>1</td>
</tr>
<tr>
<td>Number of Hidden Neurons</td>
<td>35</td>
</tr>
<tr>
<td>Number of Hidden Layers</td>
<td>1</td>
</tr>
<tr>
<td>Number of Total Inputs</td>
<td>(40x1)+(40x4)=200</td>
</tr>
<tr>
<td>Number of Cells</td>
<td>30x50=1500</td>
</tr>
<tr>
<td>Dimension of Input Matrix</td>
<td>200x1500</td>
</tr>
<tr>
<td>Network Type</td>
<td>Feed Forward Back Propagation</td>
</tr>
<tr>
<td>Pre-Processing</td>
<td>Latin-Hypercube for splitting train-test data, MinMax Scaling</td>
</tr>
<tr>
<td>Time Steps Included For Training</td>
<td>3 Time Steps (10\textsuperscript{th}, 40\textsuperscript{th} and 60\textsuperscript{th} Time Steps Out of 60 Available Time Steps)</td>
</tr>
<tr>
<td>Time Steps for Prediction</td>
<td>4 Time Steps (18\textsuperscript{th}, 34\textsuperscript{th}, 39\textsuperscript{th} and 55\textsuperscript{th} Time Steps)</td>
</tr>
</tbody>
</table>

Table 6 Details of Smart Proxy
We were also able compare the error between prediction and real values. The Fig 49 provides a visualization of predicted and real pressure at time step 546 days. The third graph in Fig. 49 shows the distribution of error in percentage along the 2D geometry of the model. Similar graphs were generated for the rest of variables at different time steps.

Figure 48. The tier system with the 4 cell in surface contact with the focal cell
Figure 49 Areal comparison between Real and Predicted Pressure along with Error (%) - 4 Time Steps (18th, 34th, 39th and 55th Time Steps)
Figure 50 Areal comparison between Real and Predicted Oil Saturation along with Error (%) - 4 Time Steps (18th, 34th, 39th and 55th Time Steps)
It was noticed that for some variables at different timescales the error between predicted and real value of variable was way above acceptable limits of error. In order to investigate the reason of such inaccuracy, histograms were generated (Fig. 47). According to our analysis, the non-normal distribution of data in training set resulted in inaccurate results. Since the training set was not capable to cover the entire range of data, the model was not trained properly. Therefore, it’s important to normalize the data before training the model.

**Fig. 51 Pressure Distribution at time step=120 Days**
4. Results/Outcomes:

The regression model applied for two different case in the last section leads us to following conclusion:

1. ROP has complex relationship with other drilling variables which cannot be explained from different empirical models alone. The data driven approach combined with statistical regression analysis provides better understanding of relationship between variables and prediction of ROP.

2. The deviation of Predicted data from real data in case of Bit Balling supports the hypothesis that there are several unknown parameters that are not the part of our data but affects the ROP significantly and may result in bit failures in some cases. Therefore it is suggested to apply multivariate analysis by using all the information available such as drilling, formation and logs.

3. Different ANN methods needs to be used separately or compared for getting better accuracy or in other cases they may be used combine to improve the accuracy of predicted ROP.

4. The multivariate regression model developed for SAGD has capability to generate results similar to reservoir simulation. The time taken by these regressions are much lesser, which provides AI an edge over other physics related mathematical models.

5. The accuracy of AI models is governed by the quality and quantity of data used for training models. Preprocessing of data before training also plays a vital role in improving the prediction accuracy. Since the data used for SAGD Models does not have normal distribution it is suggested to normalize the dataset before training the model.
References:


