2022

Hydrocarbon Pay zone Prediction using AI Neural Network Modeling.

Darren D. Guedon
dg0014@mix.wvu.edu

Hydrocarbon Pay zone Prediction using AI Neural Network Modeling.

Darren D. Guedon

A thesis submitted to 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 Science
in
Petroleum and Natural Gas Engineering

Shahab D. Mohaghegh, Ph.D., Committee Chair
Prof. Samuel Ameri, M.S, Departement Chair
Mehrdad Zamirian, Ph.D., Teaching Associate Professor

Department of Petroleum and Natural Gas Engineering

Morgantown, West Virginia
2022

Keywords: Pay zone, Artificial Intelligence, Petrophysics, Log Interpretation

Abstract

Hydrocarbon Pay zone Prediction using AI Neural Network Modeling

Darren D. Guedon

This paper captures the ability of AI neural network technology to analyze petrophysical datasets for pattern recognition and accurate prediction of the pay zone of a vertical well from the Santa Fe field in Kansas.

During this project, data from 10 completed wells in the Santa Fe field were gathered, resulting in a dataset with 25,580 records, ten predictors (logs data), and a single binary output (Yes or No) to identify the availability of Hydrocarbon over a half feet depth segment in the well. Several models composed of different predictors combinations were also tested to determine how impactful some logs were compared to others for the prediction process.

With 32 tested models using a base set of 5 logs (X, Y GR, DEPT, and CALI) and different combinations of 5 other logs (RT90, RHOB, NPHI, PE, DT). All models containing RT90, NP, or DT led to a better prediction matching the pay zone established based on a petrophysical analysis and completion data from the well.

Results from this project could be used as another support to help and justify decision-making for a Petro physicist regarding work in the field with less experience.

Acknowledgments

I would like to thank my advisor, Dr. Shahab D. Mohaghegh to introduce me to the technology of AI and machine learning through my understanding of petroleum engineering. I would also like to thank my committee members Dr. Shahab D. Mohaghegh, Dr. Mehrdad Zamirian, and Professor Samuel Ameri for their support of this thesis. A special thanks to Dr. Mehrdad Zamirian who guided me during the entire project. Thank you to both my parents for always checking on me and giving me their support during this period. Finally, I would like to thank West Virginia University and the Department of petroleum and Natural gas engineering which have been supporting me during my master's program and giving me this opportunity.

Table of Contents

List of Figures

List of Tables

Literature Review

Based on the Schlumberger oilfield glossary: "Formation Evaluation is the measurement and analysis of formation and fluid properties through examination of formation cuttings or using tools integrated into the bottomhole assembly while drilling, conveyed on wireline or drill pipe after a borehole has been drilled. It is performed to assess the quantity and producibility of fluids from a reservoir. It guides wellsite decisions, such as placement of perforations and hydraulic fracture stages, reservoir development, and production planning". Since the first electrical resistivity measurement in an oil well by the Schlumberger brothers in 1927, formation evaluation techniques and methods have vastly improved from the use of resistivity logs for water saturation determination to nuclear tools such as Gamma density to estimate shale content or Neutron log for porosity determination (Schon, 2015). In addition, a petrophysical interpretation of the measured properties play an essential role in the formation evaluation process with an early sign by Archie's famous equation (Archie 1942) describing a correlation between resistivity, water saturation, porosity, and some empirical parameters:

$$S_w = \left(\frac{R_o}{R_t}\right)^{\frac{1}{n}}$$

Nowadays, with the increase in tool logging, better and more precise petrophysical interpretations are effectuated. Results obtained from wells logs and petrophysical analyses are crucial over several paths of a petroleum engineer career. Completion engineers utilize petrophysical results to establish completion intervals; reserves engineers identify recoverable reserves and their value to their owners, asset review teams show property values for property disposal and acquisition, reservoir engineers build simulator models and depletion planning, production engineers develop and operate fields, and

enhanced oil recovery engineers planning EOR operations. All these operations are on the back of log interpretation and petrophysical analyses.

Figure 1 Formation Evaluation Operations

Among all the use and application of formation evaluation, one of interest is identifying producing zone location, also called the pay zone. In the borehole, the pay zone must be located to determine completion depth; even though properties of the reservoir play an essential role in the process, other factors such as economics, technological limitation, or restrictions can impact it. Pay zones are identified by a Petro physicist using log interpretation, each log having a role. In the case of a conventional reservoir, an ideal pay zone could be represented by a rock from a permeable and porous formation such as a sandstone with a sign of hydrocarbon. This requirement can be verified by first identifying the lithology of each formation to identify the cleans ones. This is possible using different logs, and Gamma-ray seems to be the most popular, where a low GR value suggests clean formation while high GR is related to shale content. Next, we have resistivity logs that detect permeable zones and estimate water saturation, moveable hydrocarbon, wet zone, and Hydrocarbon zone. Porosity is studied using a combination of sonic log Neutron-Density log. Following that methodology, we have a simplified interpretation of wells logs to identify pay zone.

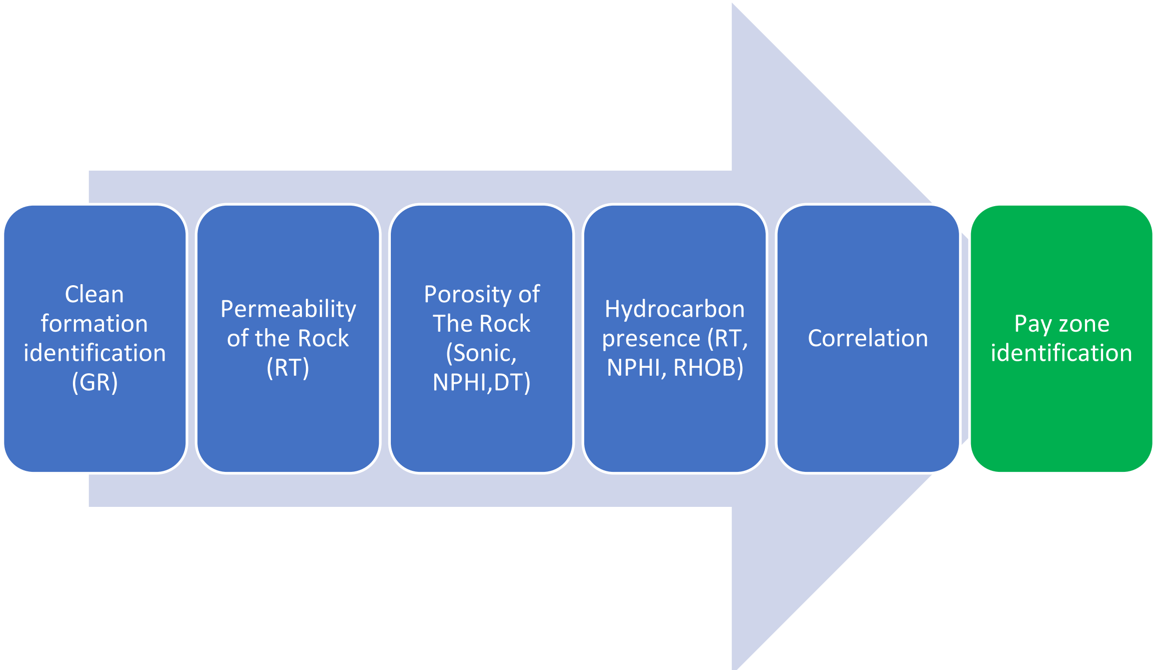

Figure 2 Pay zone identification Process

The wells logs interpretation requires significant human effort, great expertise, and experience in converting raw measurements into commercially valuable information. However, log interpretation has been made using other means via Artificial intelligence.

Considered a paradigm shift by some experts, Artificial Intelligence (AI) is the technology that mimics the human brain's cognitive abilities in analysis, modeling, and decision making. A famous example of AI is neural networks that rely on the same system used by a human brain to process information via neurons. With an estimated 10-500 billion neurons in a brain, AI neural networks are not comparable in size by only averaging in the hundreds of thousands based on the target value. However, the difference appears to be in the speed at which the information is processed, with machines being 10 million times faster than the human brain explaining why we rely more and more on them for daily tasks. This helps to introduce the sector of machine learning, which is a branch of AI whereby the use of computer algorithms is needed to learn from data instead of explicit programming. Finally, deep learning is a subset of machine learning used in neural networks by having multiple layers of neurons.

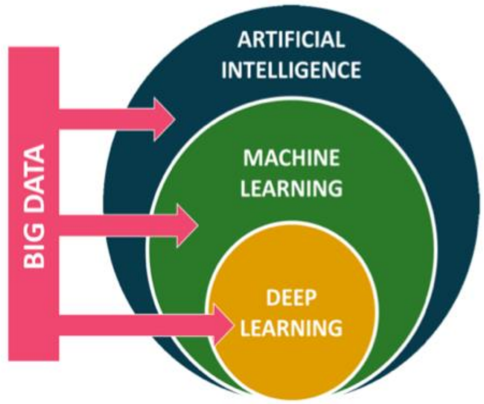

Figure 3 AI and Machine Learning relationship

Using machine learning algorithms, models can learn from the data without prior knowledge allowing feats like pattern recognition, classification, continuous value prediction, and forecasting of events or performance. To this day, three main categories of machine learning are identified: supervised learning, unsupervised learning, and reinforcement learning. The main difference between supervised and unsupervised learning is that we already know the output during supervised learning. We are trying to teach our model to identify the correct way to reach it. However, unsupervised learning lets the model find its pattern and output. Reinforcement learning is mainly used for decision-making.

Even with a significant increase in interest in machine learning in the oil and gas sector, (Figure 4), this interest is still tiny compared to other domains (Figure 5).

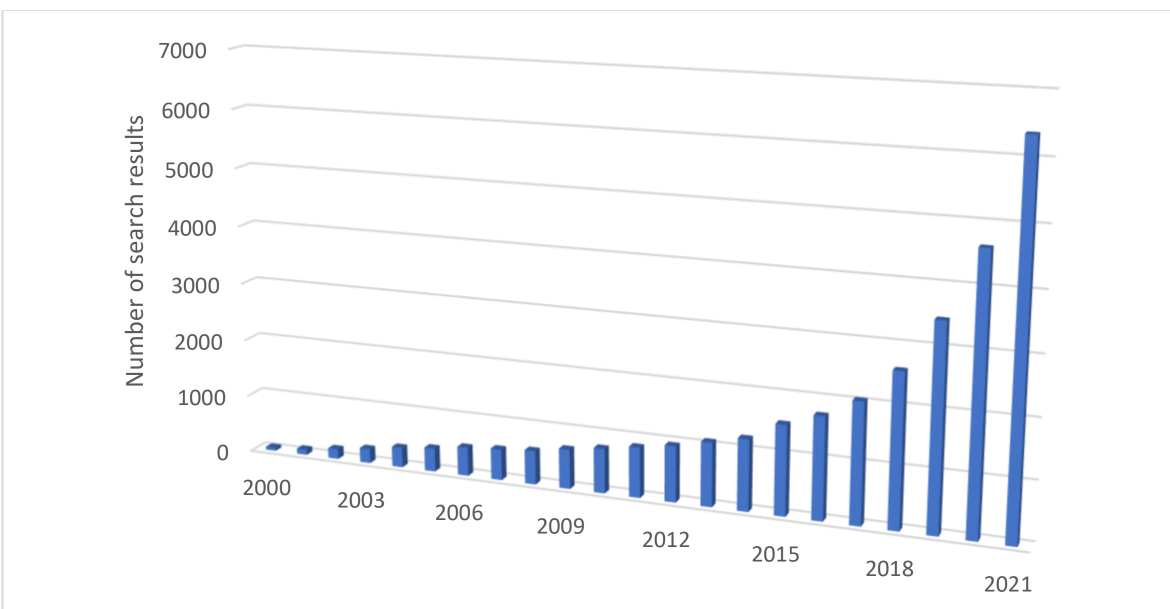

Figure 4 Google Scholar results for Petrophysics and Machine Learning since 2000

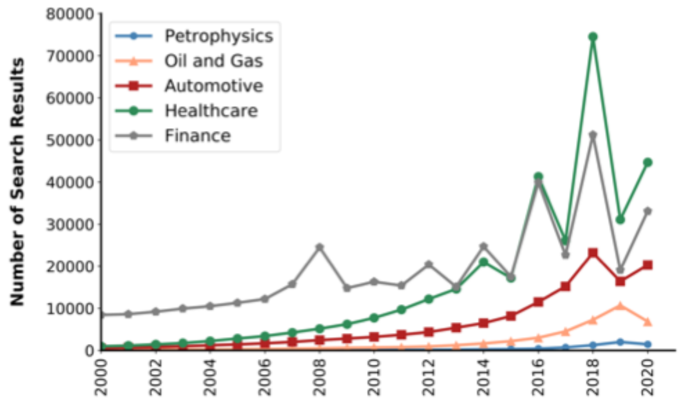

Figure 5 Google Scholar results for Machine Learning 'AND' since 2000

One exciting piece of research was done regarding supervised machine learning for automated lithology prediction from wireline log data. Over 100 wells were interpreted in five minutes during this research, yielding over 70% lithological match to the human petrophysical interpretation. The algorithm was also claimed to remove bias and inconsistencies. Some of the traditional methods are still implemented with, for example, the addition of a feature as an input to replicate the 'crossover' of neutron porosity and bulk density used by Petro physicists to separate lithology. A minimum log suite with GR, RES, NPHI, RHOB, and DRHO was also used as a based input for the model, with other logs being used if available or necessary. One of the main results of this experiment can be summarized in fig 6, illustrating three lithology predictions of the UNIVERSITY-40 well situated in the Permian Basin based on one model utilizing 4 logs as inputs: GR, neutron porosity, bulk density, and photoelectric. The first one on the second column was trained using wells from Northwest Australia (F1 score = 0.546), the third lithology column was predicted using data from the Permian Basin (F1 score = 0.826), and the fourth lithology column was predicted from data using both basins (F1 score = 0.836).

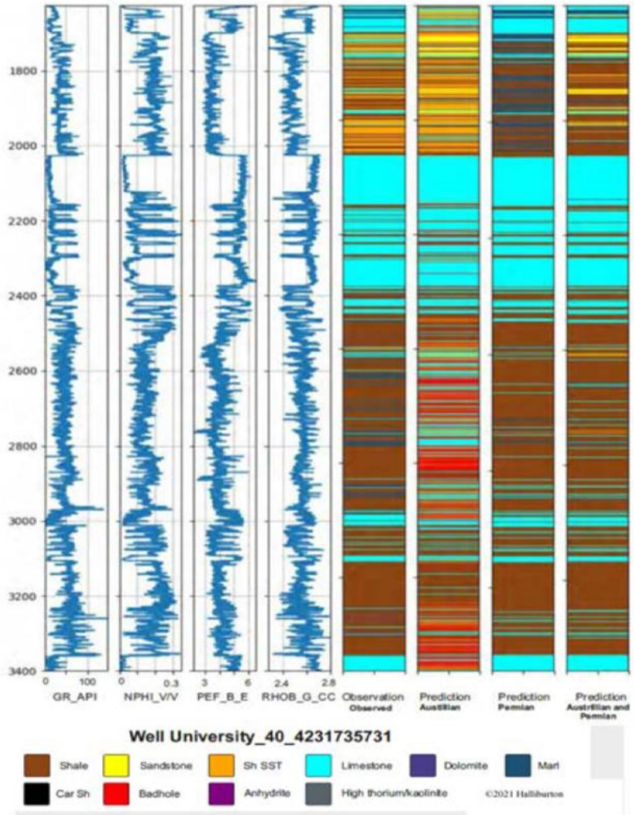

Figure 6 Lithology Prediction

Such results look promising in the petroleum industry with its wealth of data. But compared to other sectors, the oil and gas sector has been slower to integrate it as a tool for operation. Nevertheless, progress is being made, and ways to apply machine learning algorithms are implemented as described in an article discussing the use of Machine learning for gas and oil exploration. Several algorithms are being tested, such as Neural Network or Gradient Tree Boosting, leading to different results depending on the demand (Nordloh et al.). Also, machine learning model building and petrophysical projects appear to share similarities with both being time-consuming on carrying out data collation and quality checks, as described by a poll conducted on LinkedIn amongst Petro physicists and geoscientists, where a quality check can take between 50 to 90% of the time of the project (McDonald, 2021). In data analytics, poor data quality can lead to a 20% reduction in labor productivity and is the main reason for the failure of 40% of business initiatives, as Friedman and Smith (2011) mentioned. Wireline data contain missing data or outliers, which can profoundly impact the model, so it is necessary for the dataset to be pre-processed (McDonald, 2021).

Background and Problem definition

Hydrocarbon productions require drilling, completion, and maintenance expenses once the well produces. Production must be optimized by recovering the maximum hydrocarbon allowable to make such investment profitable. However, due to technological limitations, extraction is only possible at some targeted zone called pay zone, where most of the recoverable hydrocarbons are situated.

Using a Petro physicist's point of view, the pay zone can be identified using geological and log interpretation knowledge. Data collected from pilot wells or surrounding producing well give an overall understanding of the lithology of the area of interest. In addition, petrophysical interpretation allows identifying the presence of hydrocarbon. Finally, based on the analysis and interpretation made by the Petro physicist and the criteria to consider a producing zone, the pay zone is determined. Overall, this process requires the recognition of some patterns due to the nature of the reservoir and the difference in rock properties with and without hydrocarbons.

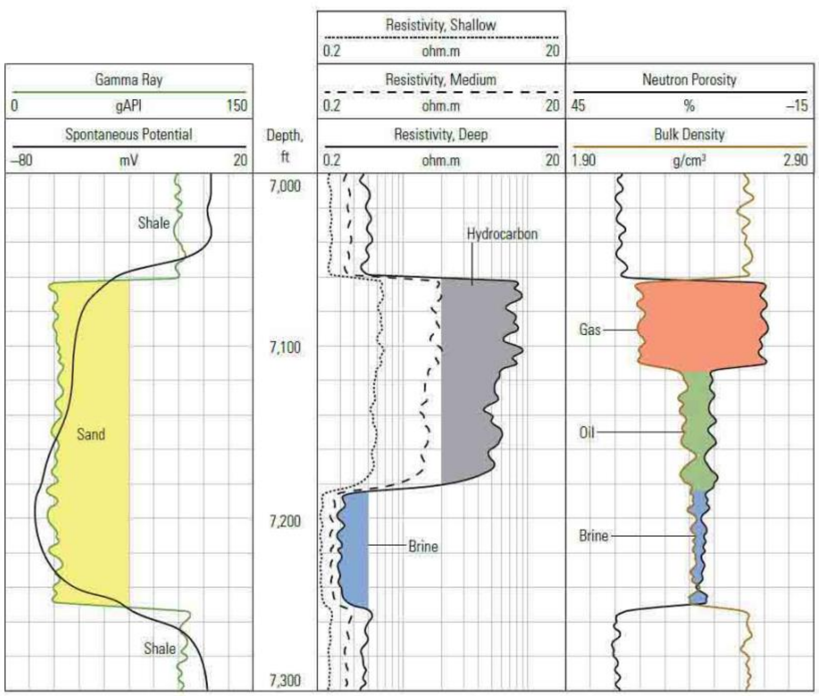

Figure 7 Pay zone identification petrophysical method

This process requires the expertise of a Petro physicist, which can be prawn to human errors. In addition, there is a degree of sensibility, such as the performance of the Petro physicist, which is impacted by different fields. This is where our study comes into interest with machine learning to complement the weakness of human interpretations. An effort has also been made to identify the minimum required logs for precise pay zone identification, which, if successful, can impact the traditional number of logs needed and expenses for formation evaluation.

Dataset Development

Despite its vast data pool, machine learning projects in the oil and gas sector are less popular due to its limitation of access. Most of the data are not public, and the ones available can be suggested to be missing information or inaccuracies. Due to the heterogeneity of the rock, it can be challenging for the model to identify the desired patterns. Spatial constraints were expected to reduce the number of wells by picking close enough ones to have similarities in their rock properties. In addition, all the wells do not follow the same set of logging, requiring the chosen wells to have enough standard logs. The listed prerequisite and limitations led to the selection of 10 verticals wells from the Santa Fe field in Kansas. Dataset development consisted first of data gathering followed by data partitioning to result in a clean dataset ready for training with precise inputs and output and means to verify the model's accuracy.

Data Gathering

The first step in the dataset development was to gather all the data and information necessary around the well of interest. Most of the wells' data were obtained from the Kansas geological survey website. Wells logs data were retrieved in the form of LAS files. The primary tool for data manipulation was the Spyder software using the Python language to build our codes. From the LAS file, all the log data at different depths were extracted and put into a comprehensive database containing the data of every well. Additional information such as top formations, completion depths, and others was also gathered.

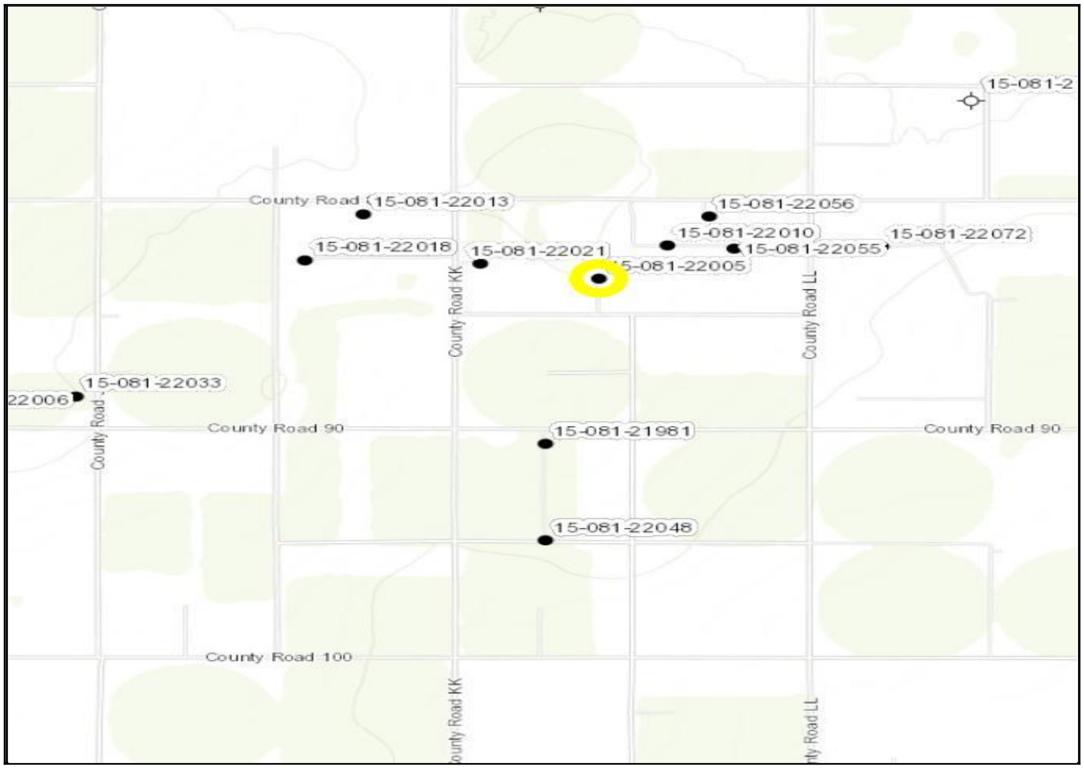

Figure 8 Wells location from Santa Fe field in Kansas

Petrophysical pay zone interpretation

During the data gathering process, completion depth was also collected. This allows us to have an idea of the existing pay zone depth that the engineering team chose to produce. However, it does not reflect all the possible pay zones. Legal restrictions, technology limitations, and budget impact the team's judgment in selecting a pay zone. Since this project focus on the capability of an AI model to identify a pay zone base on the rock properties, the depth of the petrophysical pay zone will be more fitted as the main objective to predict rather than the depth established by the completed pay zone. Petrophysical pay zone can be identified using the software PETRA, and the log data gathered.

The first step for pay zone interpretation using PETRA is implementing logs into tracks for better visualization. This also allows the logs to be grouped with respect to the rock characteristics we are looking for. Figure 9 is a snippet of the log track format we chose for our interpretation. Five main tracks were used. In the first one, we grouped Mud resistivity, Caliper, SP, and Gamma-ray mainly to identify clean rocks (Sandstone, limestone) from non-cleaned rock (Shale). SP can be seen as a deflection from the shale line, with a higher deflection representing more permeability. In addition, deflection in the caliper line is suggested to mud cake, a sign of a permeable zone. Finally, a lower gamma-ray value than the average shale line is a sign of clean rock. We put all the Resistivity logs on the second track to correlate shallow and deep resistivity. Permeable zones are characterized by higher resistivity deeper in the formation than shallower zones. The opposite is seen in a non-permeable zone. The third track is composed of neutron-density logs for identifying the porous area. It also helps in the rock lithology identification and the nature of the fluid in it. The fourth and fifth tracks were used for density correction and sonic log, respectively, serving as confirmative tools from previous interpretations made using other logs. Not all the logs used for track visualization were used as input for model prediction showing the ability of machine learning to achieve a similar result with fewer data.

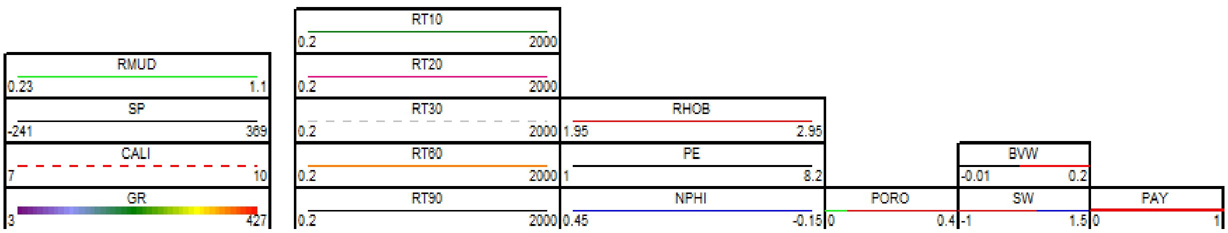

Figure 9 PETRA log track format

Once the track model is completed, we can add a limitation to differentiate the possible hydrocarbon-bearing zone from to pay zone. This is where the experience of the Petro physicist in the field can be used. To distinguish the clean zone from the dirty zone, we proceed to calculate the shale volume of the entire log depth and then consider as clean a shale volume of less than 20% using the following formula:

$$V_{sh} = I = \frac{GR_{Value} - GR_{min}}{GR_{max} - GR_{min}} * 100$$

Among the clean zone we identified the permeable area by using a cut-off with average porosity above 7% and by looking at the separation between Deep resistivity and Shallow resistivity:

$$\emptyset = \frac{\emptyset N + \emptyset D}{2}$$

Once we identify a clean water-bearing zone, we proceed to find Rw by recording the Sp from the area of choice. Assuming that Rw was constant the whole log, we were able to use Rt data and the average porosity to find the water saturation and the bulk volume of water (BVW), such as:

$$BVW = \emptyset * Sw$$

Using the cut-off value of Sw=50%, porosity>=7%, BVW<0.09, Thickness at 2 ft and a=0.80, m=n=2, we determined the petrophysical pay zone.

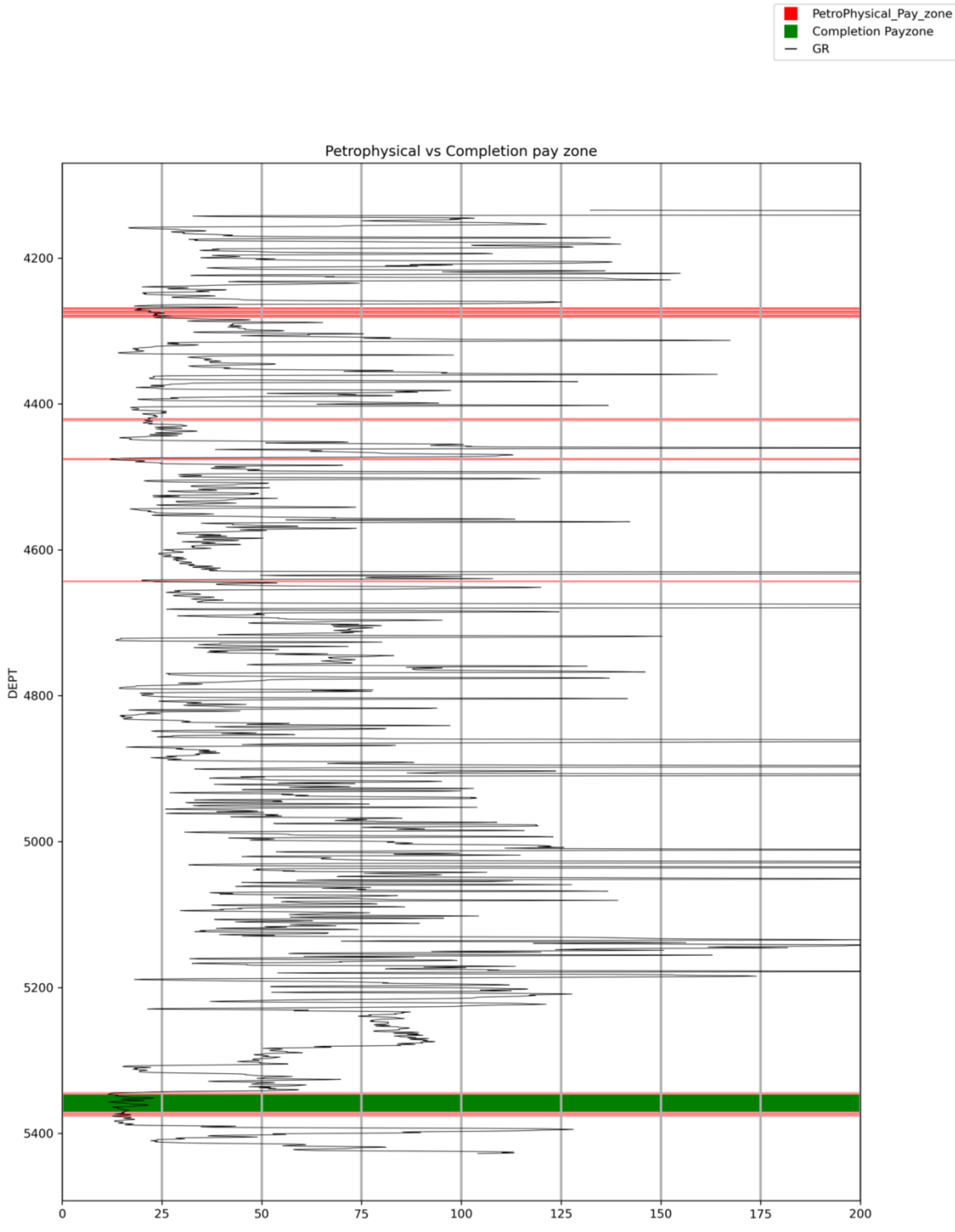

Figure 10 Petrophysical vs. completion pay zone

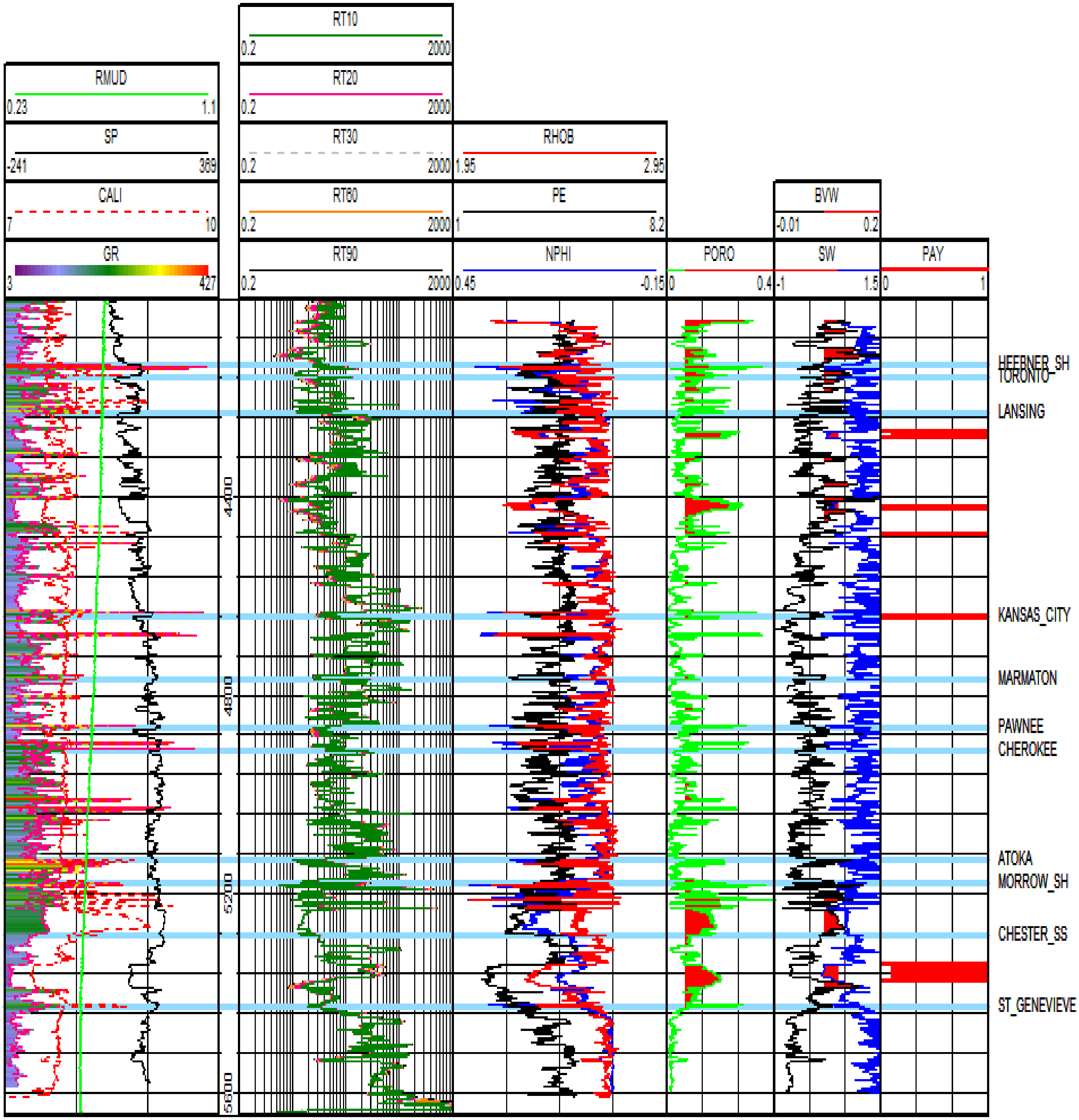

Figure 11 Petrophysical interpretation on blind WELL

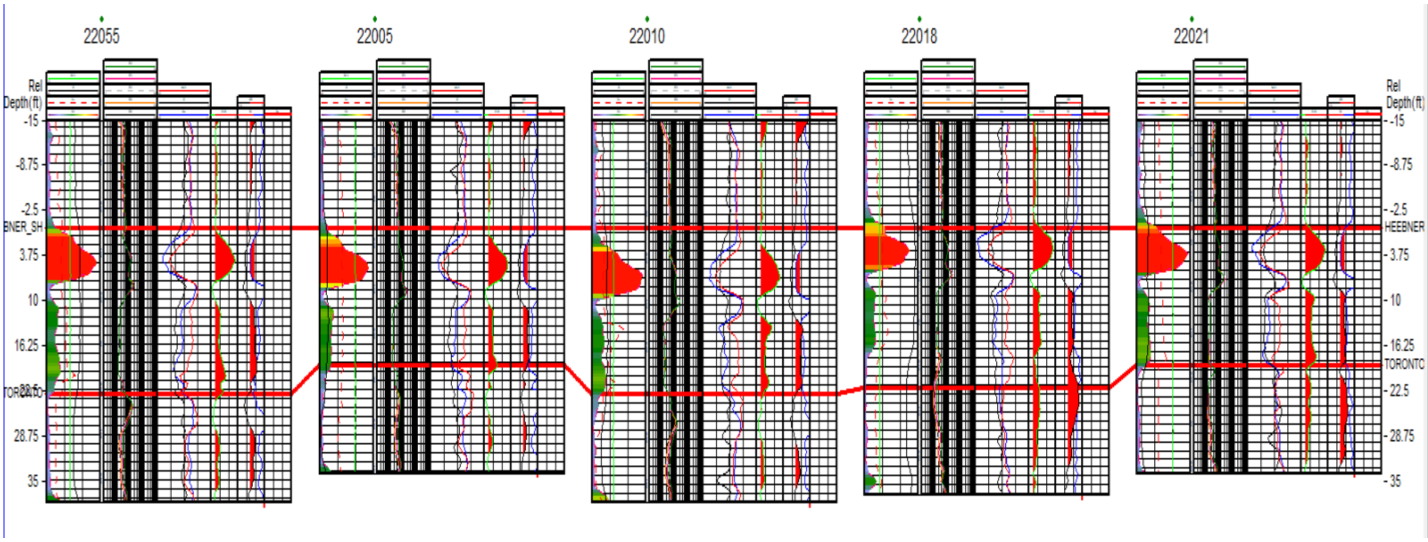

Figure 12 Petrophysical interpretation (HEEBNER - TORONTO top formation).

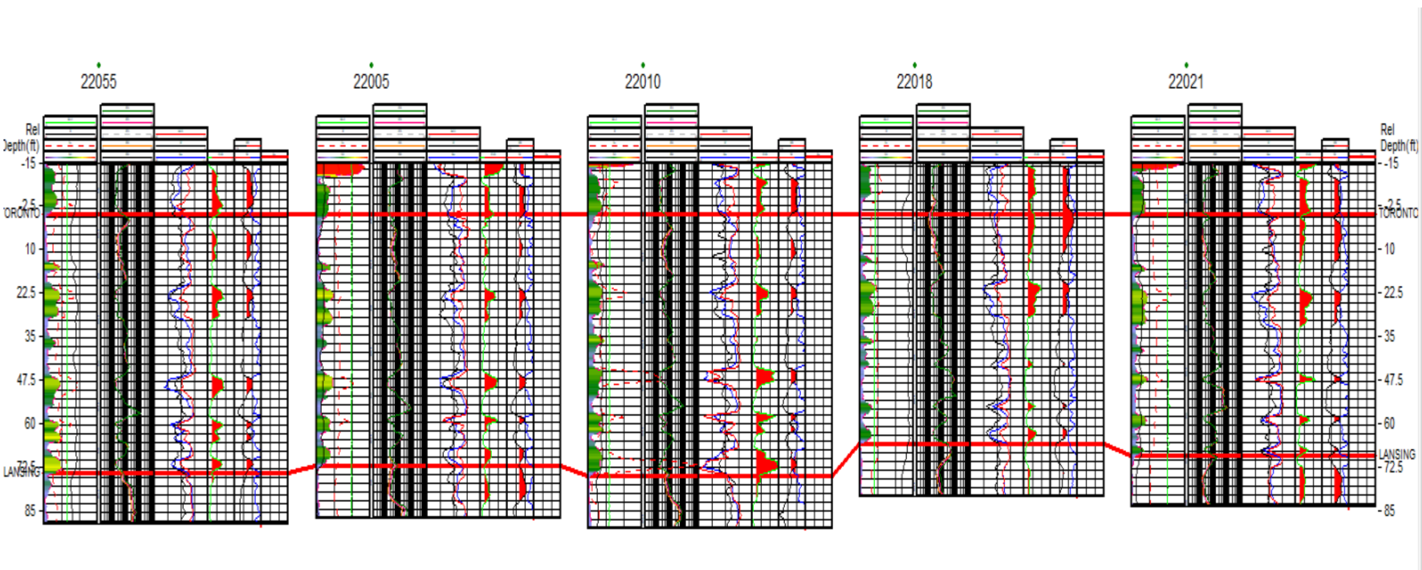

Figure 13 Petrophysical interpretation (TORONTO- LANSING top formation).

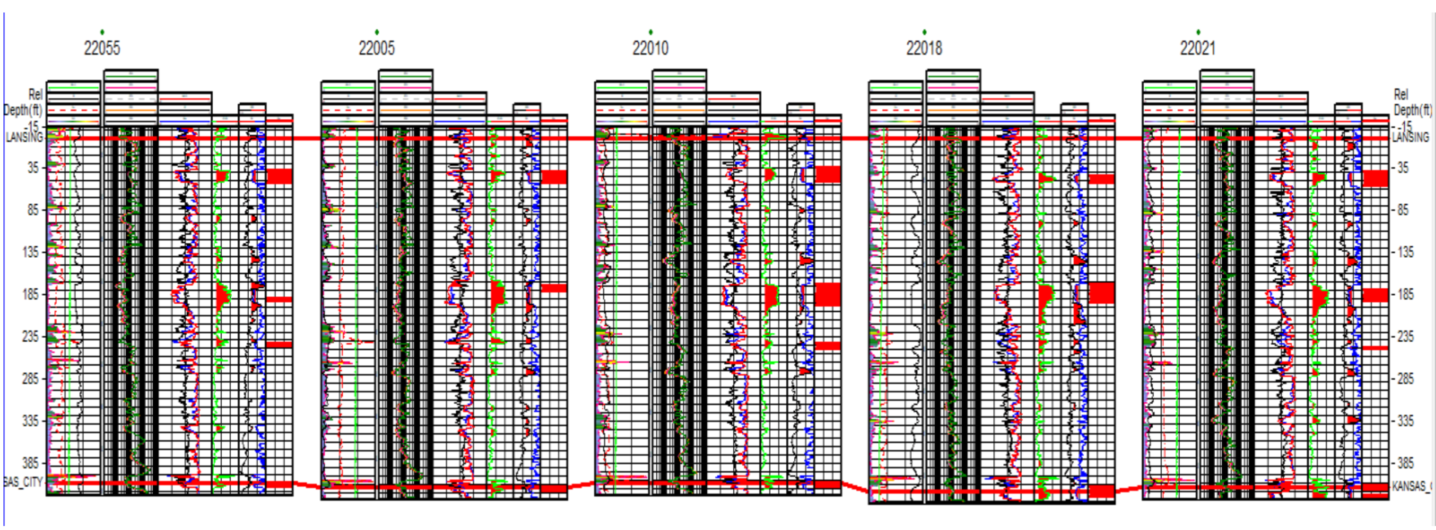

Figure 14 Petrophysical interpretation (LANSING - KANSAS CITY top formation).

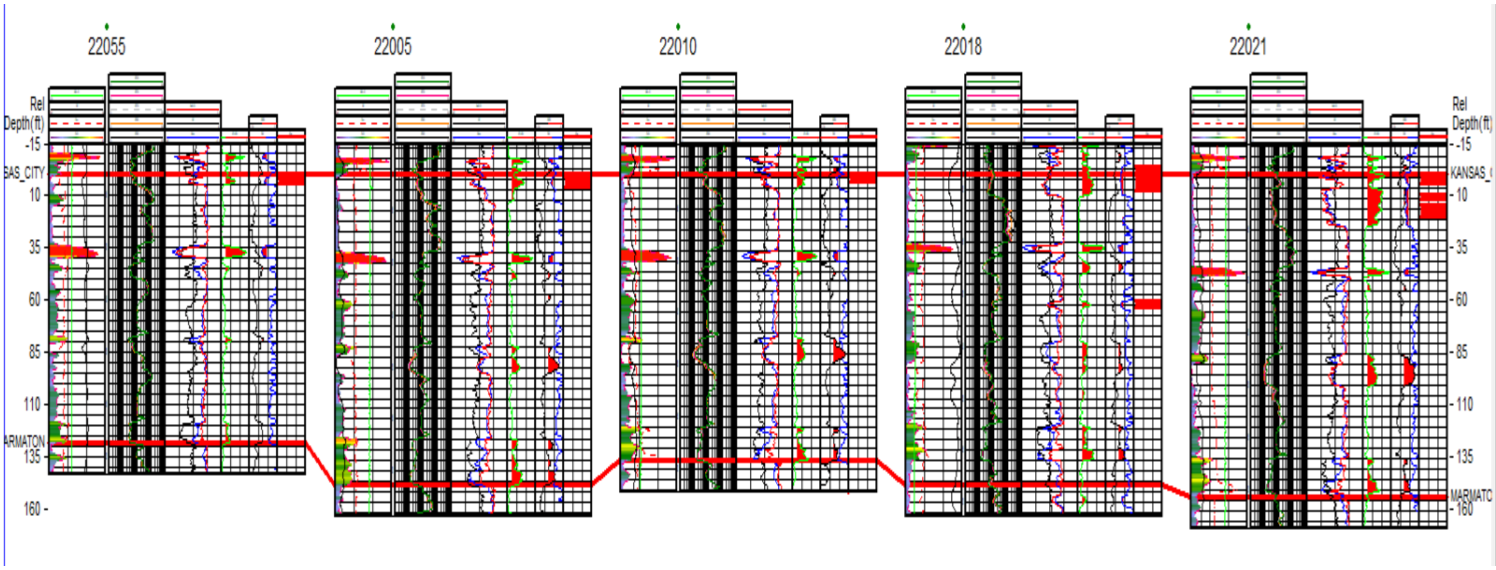

Figure 15 Petrophysical interpretation (KANSAS CITY - MARMATON top formation).

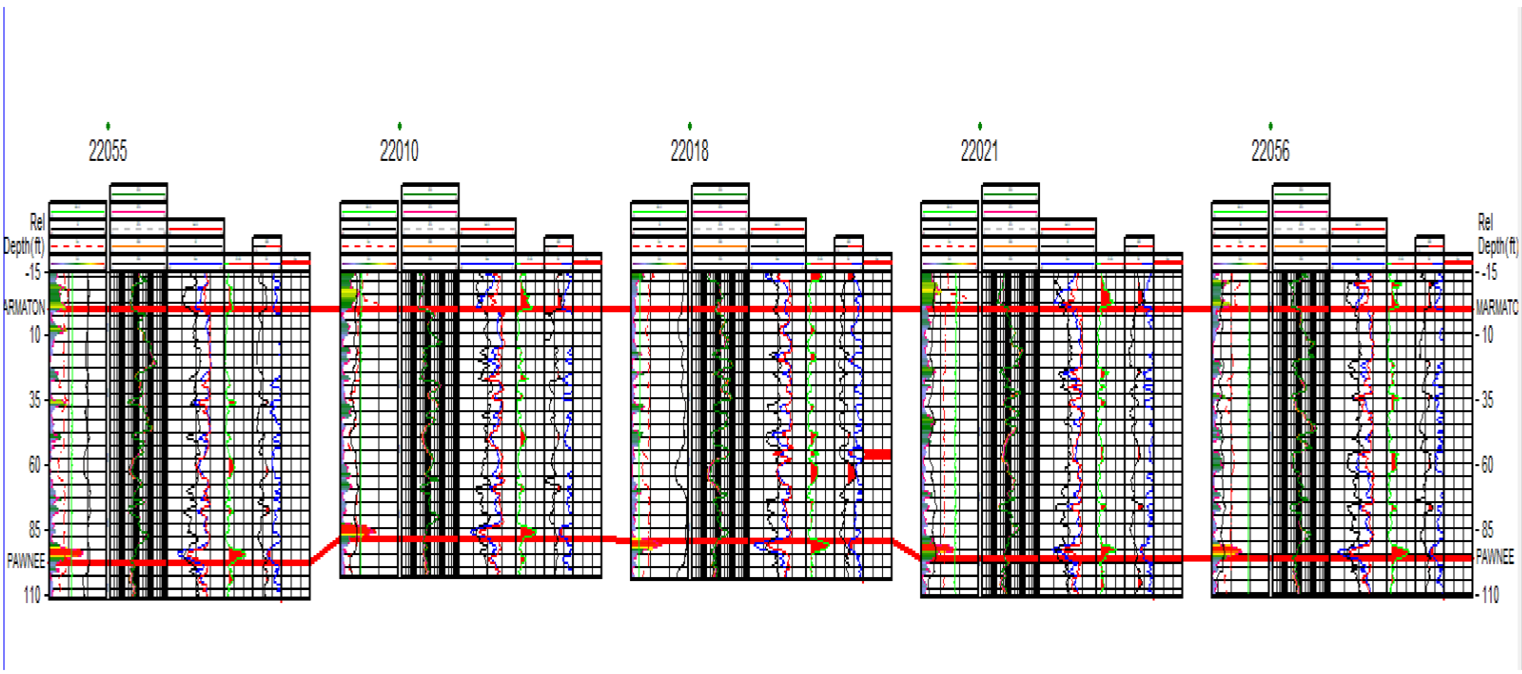

Figure 16 Petrophysical interpretation (MARMATON - PAWNEE top formation).

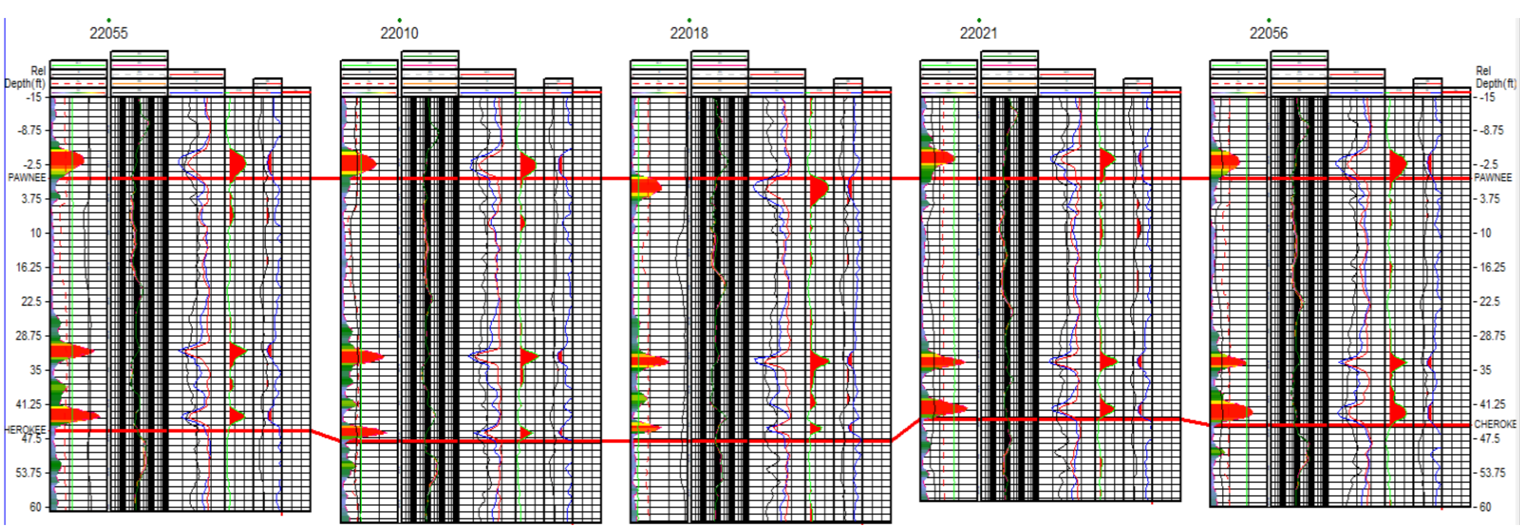

Figure 17 Petrophysical interpretation (PAWNEE - CHEROKEE top formation).

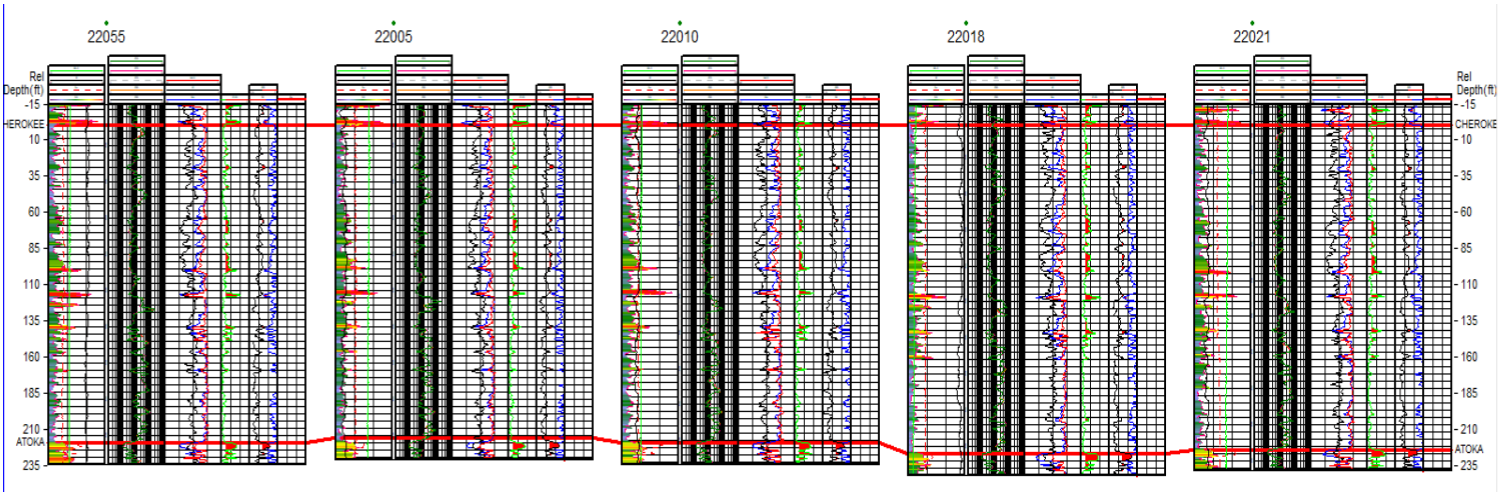

Figure 18 Petrophysical interpretation (CHEROKEE - ATOKA top formation).

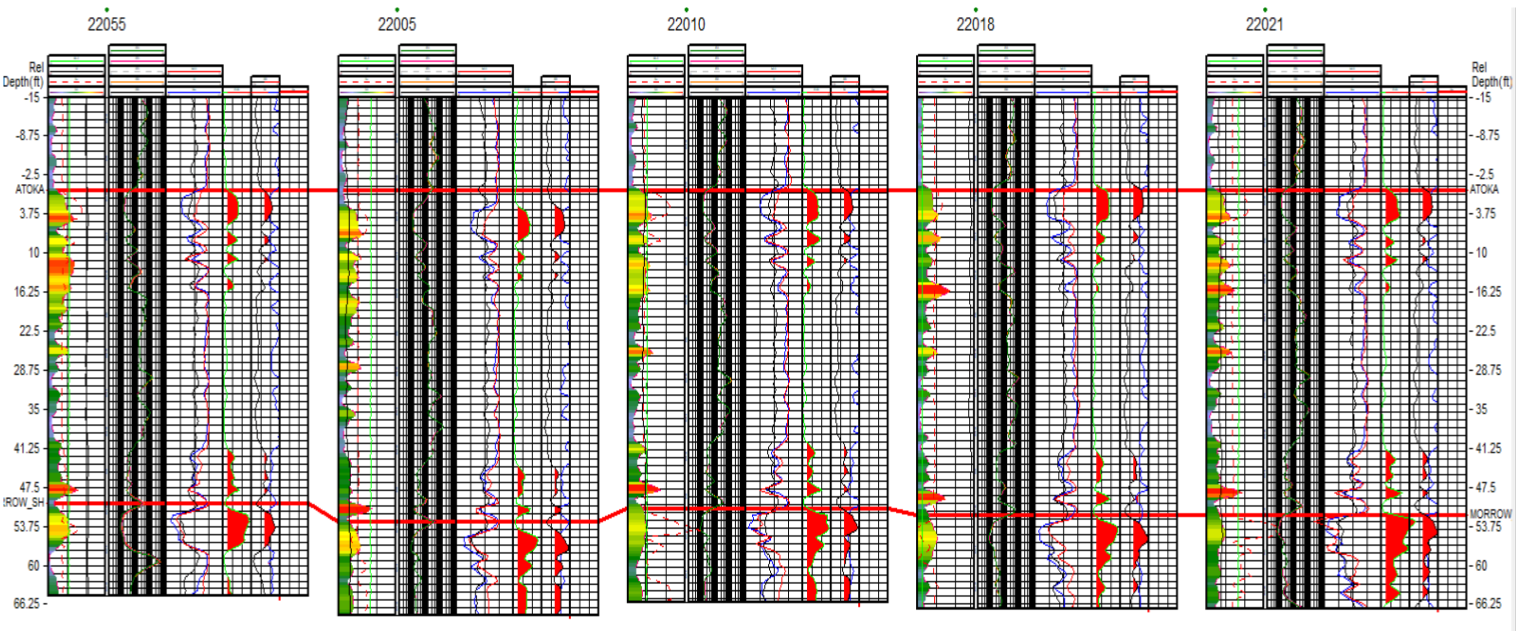

Figure 19 Petrophysical interpretation (ATOKA - MORROW top formation).

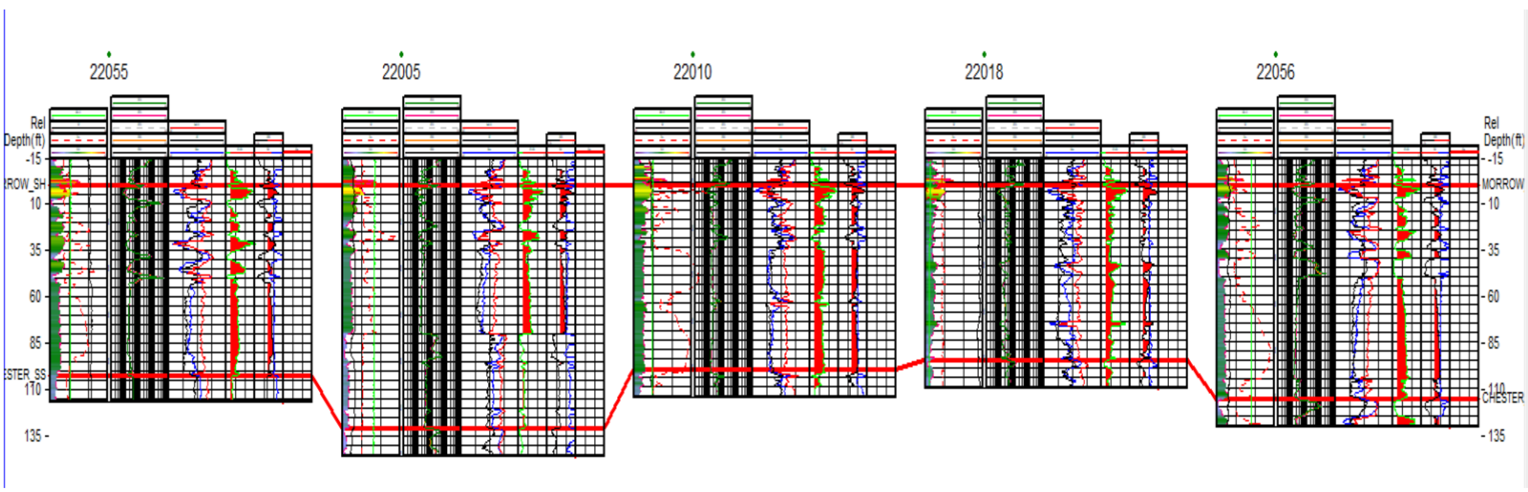

Figure 20 Petrophysical interpretation (MORROW - CHESTER top formation).

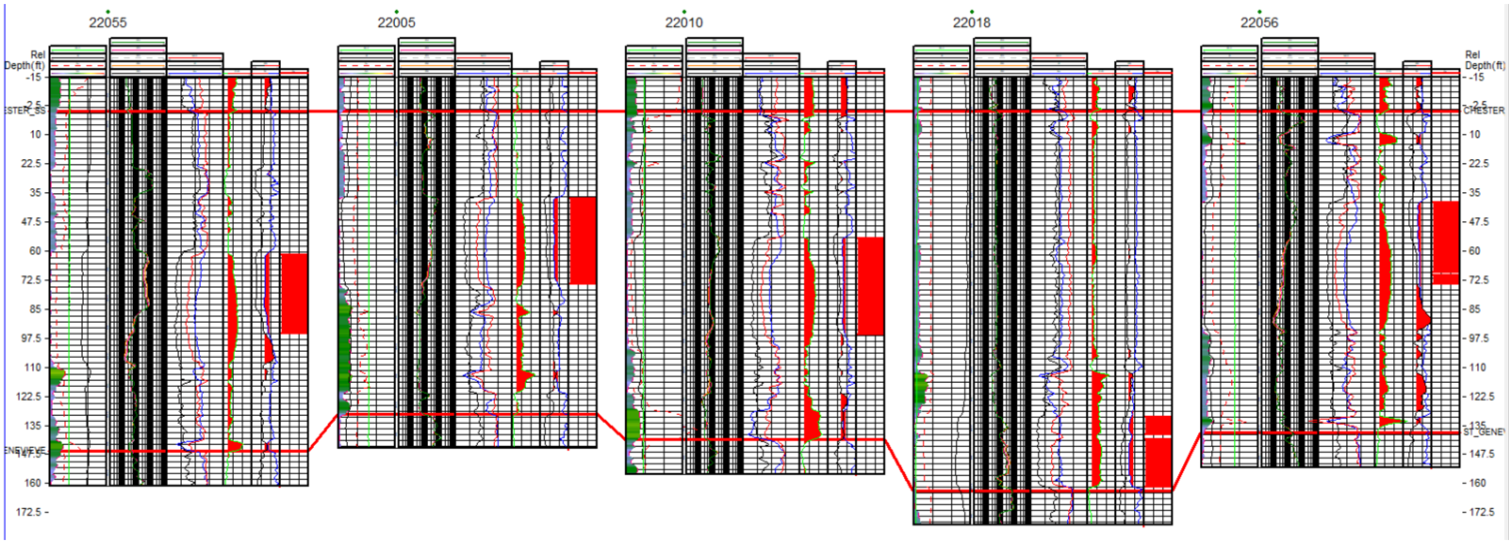

Figure 21 Petrophysical interpretation (CHESTER - ST GENEVIEVE top formation).

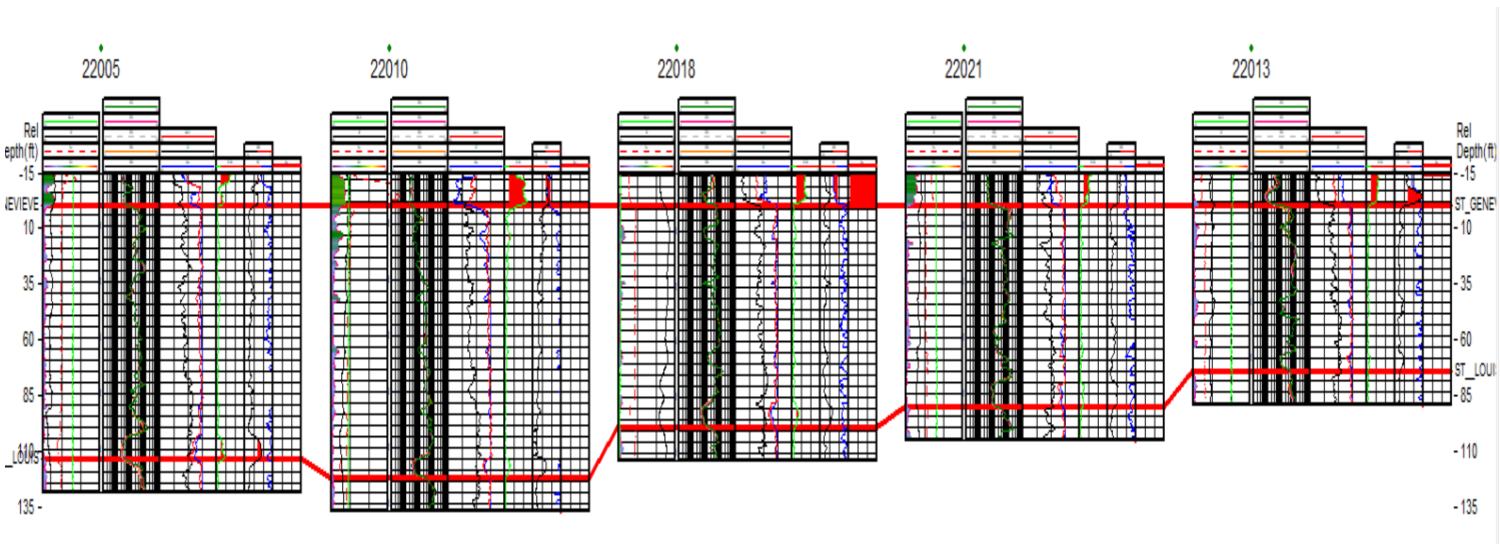

Figure 22 Petrophysical interpretation (ST GENEVIEVE - ST LOUIS top formation).

Data Pre-processing

Once the database was obtained, more pre-processing was required before getting a dataset ready for training. This step is like the data pre-processing effectuated by Petro physicist before interpretation consisted of removing all the logs that were not shared between all the wells. With 42 available logs, only five were considered essential for our base model. Highlighted in green in Table 1, it consisted of the X and Y location of each well, Depth (DEPT), caliper (CALI), and Gamma-Ray (GR) noted that these are consistent logs that are always effectuated. Next, the logs highlighted in yellow were used in different combinations to understand their impact and improve the model's performance. Finally, in red and grey are the logs that either were not available in every well (Gray) or were just a representation of the same property of the rock, such as the different resistivity or did not have a tangible impact on the model toward the identification of a pay zone.

22018	22021	21981	22055	22048	22013	22006	22010	22005	22056
Loc X	Loc X	Loc X	Loc X	Loc X	Loc X	Loc X	Loc X	Loc X	Loc X
Loc Y	Loc Y	Loc Y	Loc Y	Loc Y	Loc Y	Loc Y	Loc Y	Loc Y	Loc Y
CALI	CALI	CALI	CALI	CALI	CALI	CALI	CALI	CALI	CALI
CT90	CT90	CT90	CT90	CT90	CT90	CT90	CT90	CT90	NA
DELTAT	DELTAT	DELTAT	DT	DT	DT	DT	DELTAT	DELTAT	DT
DEPT	DEPT	DEPT	DEPT	DEPT	DEPT	DEPT	DEPT	DEPT	DEPT
DLIM	DLIM	DLIM	DLIM	DLIM	DLIM	DLIM	DLIM	DLIM	DLIM
DPHD	DPHD	DPHD	DPHD	DPHD	DPHD	DPHD	DPHD	DPHD	DPHD
DPHI	DPHI	DPHI	DPHI	DPHI	DPHI	DPHI	DPHI	DPHI	DPHI
DPHS	DPHS	DPHS	DPHS	DPHS	DPHS	DPHS	DPHS	DPHS	DPHS
DRHO	DRHO	DRHO	DRHO	DRHO	DRHO	DRHO	DRHO	DRHO	DRHO
GR	GR	GR	GR	GR	GR	GR	GR	GR	GR
ITTT	ITTT	ITTT	ITTT	ITTT	ITTT	ITTT	ITTT	ITTT	ITTT
MINV	MINV	MINV	MINV	MINV	NPHD	MINV	MINV	MINV	MINV
MNOR	MNOR	MNOR	MNOR	MNOR		MNOR	MNOR	MNOR	MNOR
NPHD	NPHD	NPHD	NPHD	NPHD		NPHD	NPHD	NPHD	NPHD
NPHI	NPHI	NPHI	NPHI	NPHI	NPHI	NPHI	NPHI	NPHI	NPHI
NPHL	NPHL	NPHL	NPHL	NPHL	NPHL	NPHL	NPHL	NPHL	NPHL
NPHS	NPHS	NPHS	NPHS	NPHS	NPHS	NPHS	NPHS	NPHS	NPHS
PE	PE	PE	PE	PE	PE	PE	PE	PE	PE
QF	QF	QF	QF	QF	QF	QF	QF	QF	QF
QN	QN	QN	QN	QN	QN	QN	QN	QN	QN
RHOB	RHOB	RHOB	RHOB	RHOB	RHOB	RHOB	RHOB	RHOB	RHOB
RMUD	RMUD	RMUD	RMUD	RMUD	RMUD	RMUD	RMUD	RMUD	RMUD
RT	RT	RT	RT	RT	RT	RT	RT	RT	RT
RT10	RT10	RT10	RT10	RT10	RT10	RT10	RT10	RT10	RT10
RT20	RT20	RT20	RT20	RT20	RT20	RT20	RT20	RT20	RT20
RT30	RT30	RT30	RT30	RT30	RT30	RT30	RT30	RT30	RT30
RT60	RT60	RT60	RT60	RT60	RT60	RT60	RT60	RT60	RT60
RT90	RT90	RT90	RT90	RT90	RT90	RT90	RT90	RT90	RT90
RXO	RXO	RXO	RXO	RXO	RXO	RXO	RXO	RXO	RXO
RXRT	RXRT	RXRT	RXRT	RXRT	RXRT	RXRT	RXRT	RXRT	RXRT
SP	SP	SP	SP	SP	SP	SP	SP	SP	SP
SPHI	SPHI	SPHI	SPHI	SPHI	SPHI	SPHI	SPHI	SPHI	SPHI
TENS	TENS	TENS	TENS	TENS	TENS	TENS	TENS	TENS	TENS
NA	NA	NA	NA	NA	NA	NA	NA	NA	SPHD
NA	NA	NA	NA	NA	NA	NA	NA	NA	SPHL
NA	NA	NA	NA	NA	NA	NA	NA	NA	SPHS
NA	NA	NA	NA	NA	NA	NA	NA	NA	AHVT
NA	NA	NA	NA	NA	NA	NA	NA	NA	BHVT
NA	NA	NA	NA	NA	NA	NA	NA	NA	CORM
NA	NA	NA	NA	NA	NA	NA	NA	NA	CORP

Table 1 LOG representation in each well

Next, depths with missing logs values were also removed. This resulted in the creation of a clean dataset with 28,152 records.

Model Development

With the current dataset, 32 models were developed using inputs based on 5 logs (X, Y GR, DEPT, and CALI) and a different combination of 5 other logs (RT90, RHOB, NPHI, PE, and DT).

The first step consisted of identifying each log's minimum and maximum values to include them in the training data. Models tend to have a higher rate of misprediction while trying to forecast value outside the range of training. This step also allowed us to find a blind well in which the data will not be used in training but to verify the accuracy of our model. Table 2 resumes the previous step leading to well 22055 being chosen as a blind well with 2572 records leaving 25,580 records for training.

	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min
Well_API	21981	21981	22005	22005	22006	22006	22010	22010	22013	22013	22018	22018	22021	22021	22048	22048	22055	22055	22056	22056
DEPT	5658.5	4069	5658.5	4060.5	5658.5	4060.5	5658.5	4100	5658.5	4060.5	5595	4060.5	5658.5	4095	5658.5	4060.5	5582	4060.5	5516	4060.5
RHOB	2.7419	2.144	2.7484	2.1511	2.7168	1.8855	2.7539	1.9599	2.7261	2.1471	2.7318	2.1205	2.7295	1.7586	2.7413	1.5898	2.7355	2.0227	2.7408	1.7722
PE	5.4555	2.0917	5.679	1.9196	5.5226	2.0996	5.9125	1.9604	6.1163	1.6677	5.5156	1.6969	5.5777	2.2548	5.3678	1.8352	5.5499	1.9333	5.3336	1.9924
NPHI	0.423	-0.0003	0.3929	-0.0056	0.3779	-0.0047	0.427	-0.0053	0.4237	-0.0053	0.4629	-0.0011	0.447	-0.0028	0.4549	-0.0028	0.4103	-0.0034	0.422	-0.0004
CALI	8.8852	7.2365	10.0142	7.7243	8.9436	7.5749	13.3086	7.2138	8.5705	7.1539	8.3759	7.0687	12.0753	7.9892	11.8141	7.3248	9.3621	7.3466	9.8863	7.4522
RT90	2000	0.5405	2000	0.9467	744.6948	1.0693	1348.667	0.9598	2000	0.7651	697.951	0.8396	2000	0.9046	938.3542	0.6077	1005.141	0.8832	411.6687	0.8702
RT60	2000	0.5557	2000	1.0082	630.1741	1.1099	1127.347	1.0268	2000	0.8035	376.4178	0.8907	2000	0.9419	615.598	0.639	945.9243	0.931	407.8517	0.9251
RT30	2000	0.5646	2000	1.0446	611.2881	1.1602	831.2721	1.1096	2000	0.8547	310.1852	0.9574	2000	0.9774	762.6137	0.6714	868.9893	0.9902	420.8004	0.9997
RT20	2000	0.5989	2000	1.1271	595.0369	1.2782	625.1349	1.1383	2000	0.9713	285.6312	1.1235	2000	1.0732	586.1022	0.7589	773.0101	1.1485	389.6828	1.1882
RT10	2000	0.6559	2000	1.1753	724.2252	1.2783	458.9001	1.1299	2000	0.9956	272.5736	1.0382	2000	0.9746	720.0075	0.9315	702.3134	1.099	462.2809	1.5876
GR	426.7371	14.343	416.6765	9.7079	338.4851	9.0215	413.1443	15.3363	359.296	8.9023	367.5644	11.1379	416.9745	9.3396	393.719	13.6467	405.5332	11.5412	355.0775	9.7837
DT	120.1559	43.1208	123.96	42.5164	121.9187	43.1917	114.3311	43.0902	139.1623	41.7624	134.2872	43.2917	128.5747	42.6145	129.2638	44.1512	114.0256	43.7805	130.1059	41.229

Table 2 Min and Max of log value per well

Since it is a supervised learning experiment, the pay zone location must be fed to the model to be able to correct itself in cases of mispredictions. This was possible by gathering data from the well's completion, looking at traditional petrophysical interpretation, and using other established software such as PETRA from which depth of potential pay zone can be retrieved.

With the input data defined and the output being a depth at which a pay zone could be located, the model was based on the scikit learn library using its classifier algorithm. The following table gives an idea of the hyperparameters used.

Hidden Layer size	(100,)	Batch size	50
Activation Function	Relu	Max iteration	10000
Initial Learning Rate	0.001	Tolerance	1e-8
iteration with no changes	100	Validation Fraction	0.1
Random State	42	Early Stopping	True
Solver	Adam		

Table 3 Model Hyper Parameters

Results

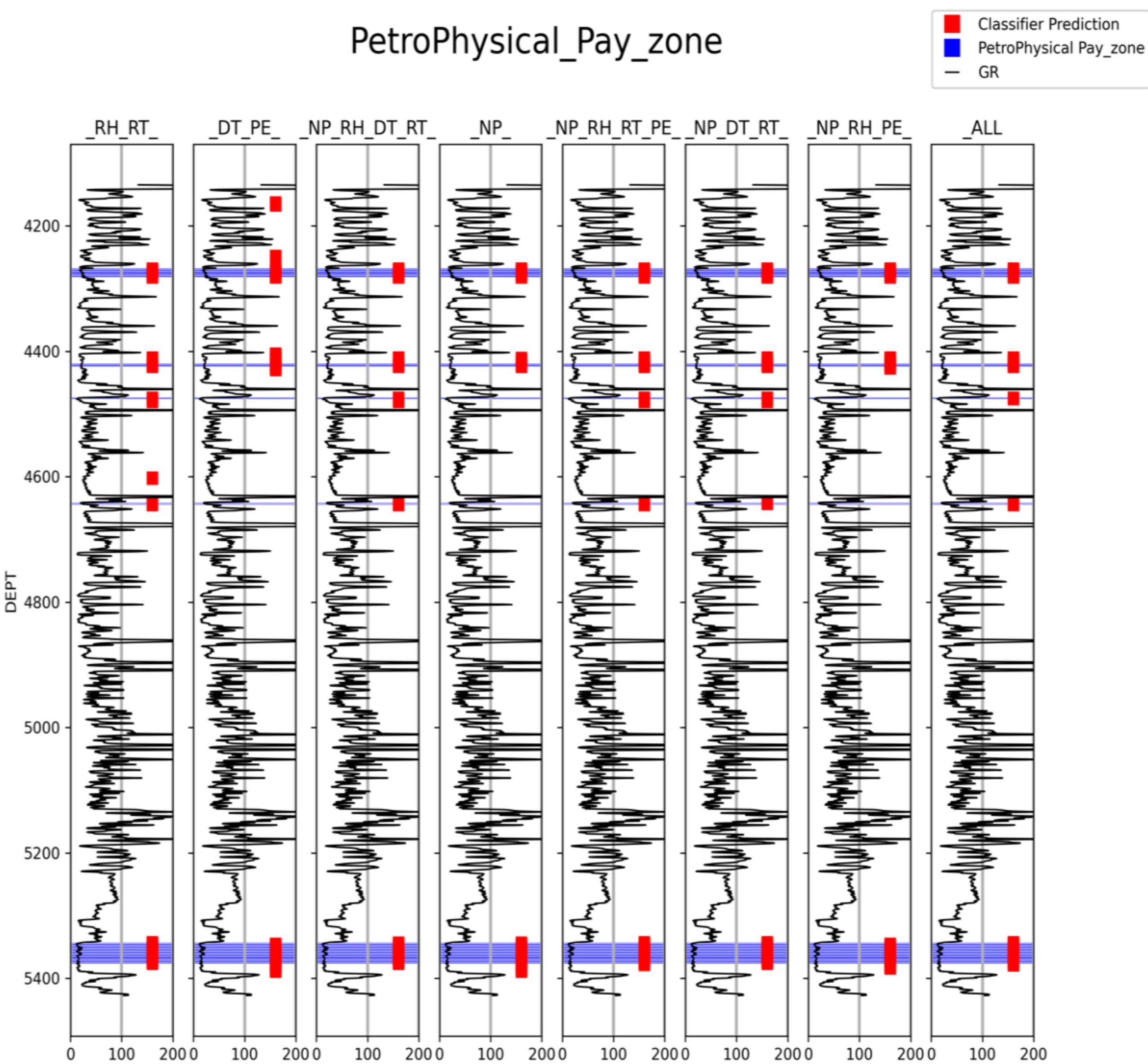

Figure 23 Pay zone prediction basic log included (X, Y GR, DEPT, and CALI) Part 1

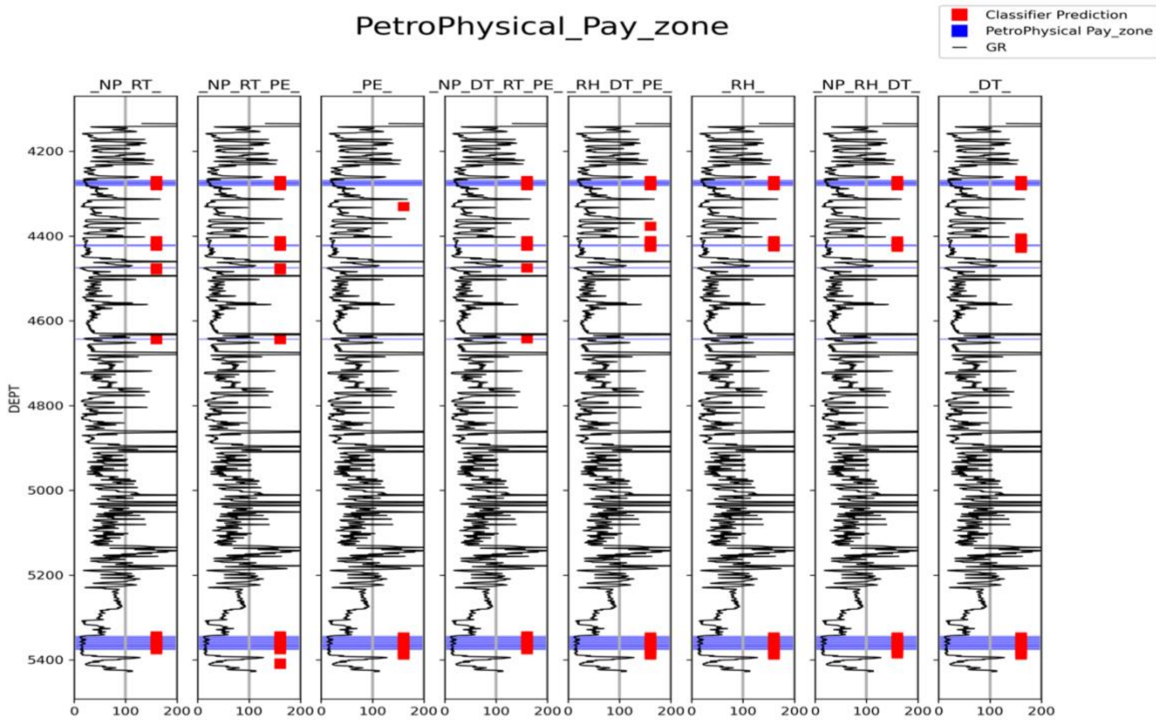

Figure 24 Pay zone prediction basic log included (X, Y GR, DEPT, and CALI) Part 2

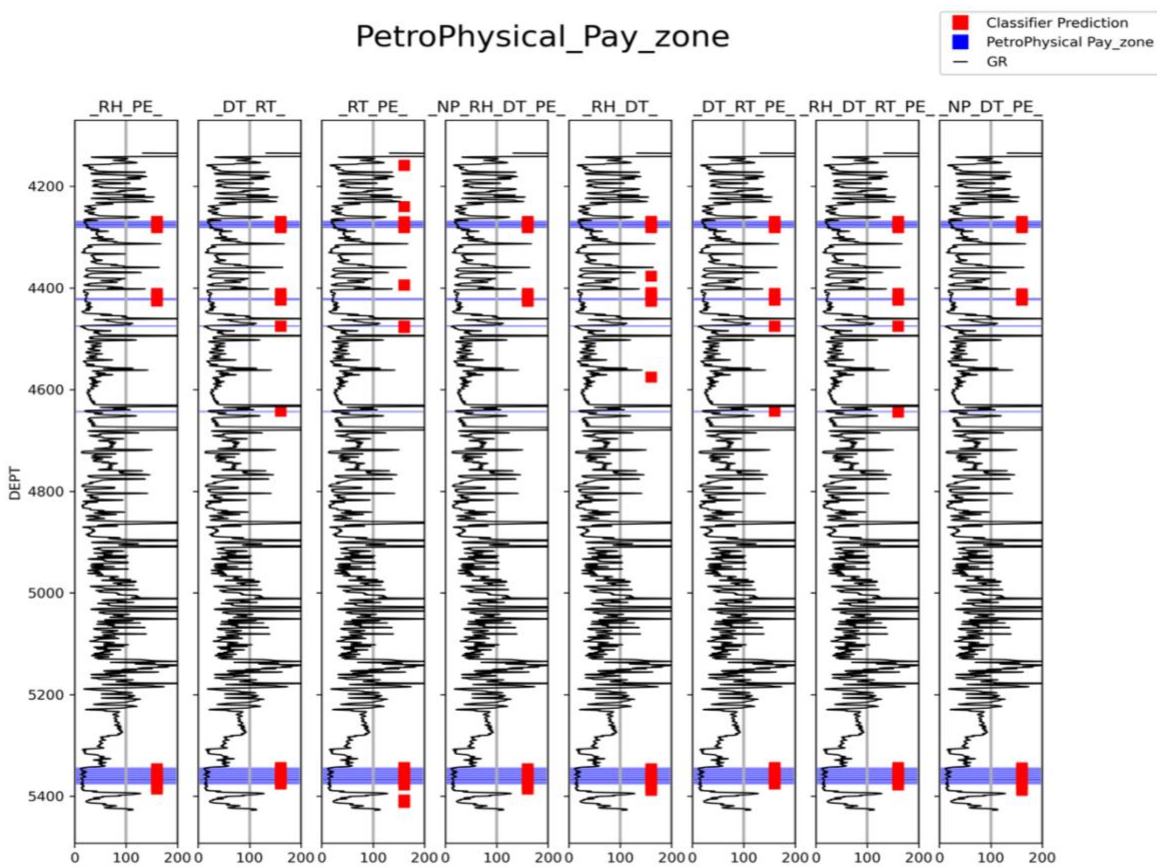

Figure 25 Pay zone prediction basic log included (X, Y GR, DEPT, and CALI) Part 3

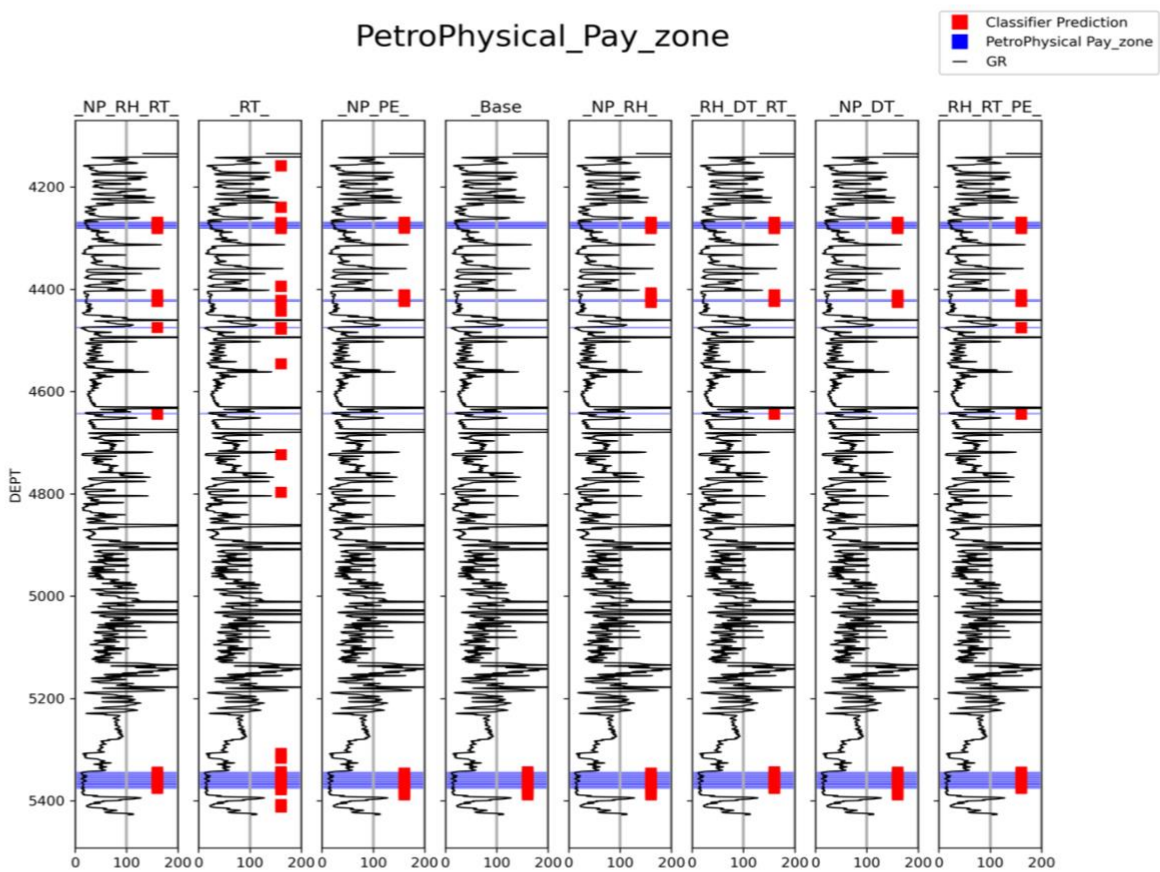

Figure 26 Pay zone prediction basic log included (X, Y GR, DEPT, and CALI) Part 4

Model's name list			
1	NP_DT_RT_PE	17	RH_DT
2	NP_RH_DT_RT	18	RH_RT
3	NP_DT_RT_PE	19	RH_PE
4	NP_RH_RT_PE	20	NP_RT
5	NP_RH_DT_PE	21	DT_RT
6	RH_DT_RT_PE	22	NP_DT
7	NP_RH_PE	23	NP_RH
8	NP_RH_DT	24	NP_PE
8	NP_RH_RT	25	DT_PE
9	NP_DT_PE	26	RT_PE
10	NP_RT_PE	27	RT
11	RH_RT_PE	28	DT
12	RH_DT_PE	29	RH
13	RH_DT_RT	30	PE
14	DT_RT_PE	31	ALL
16	NP_DT_RT	32	BASE

Table 4 List of models

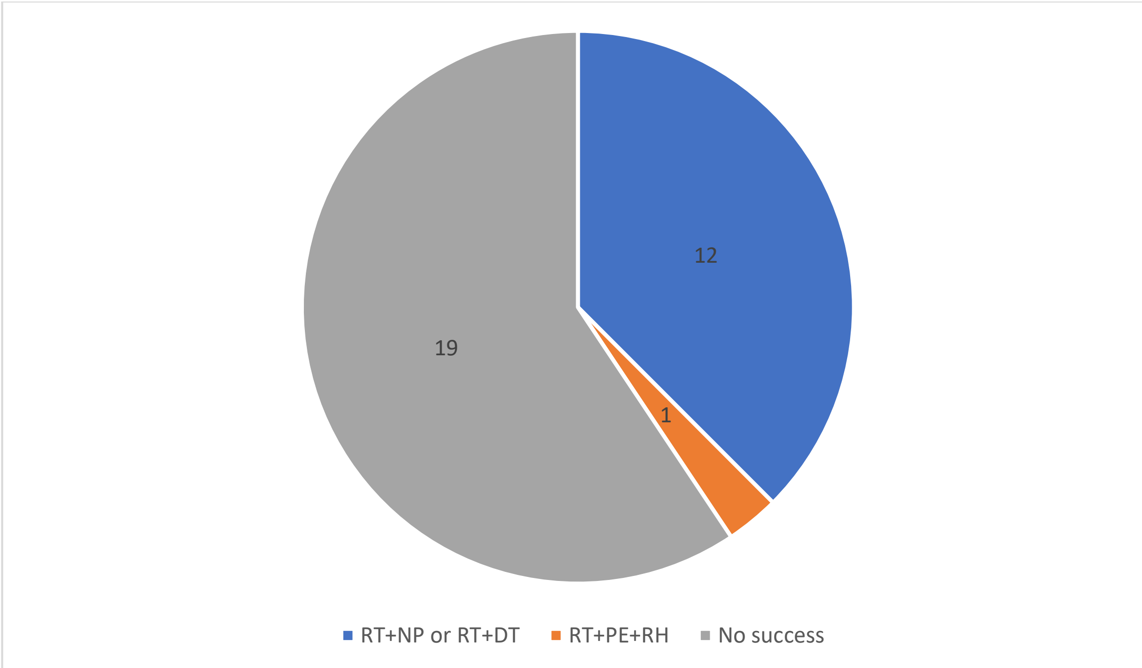

Figure 27 Model accuracy repartition

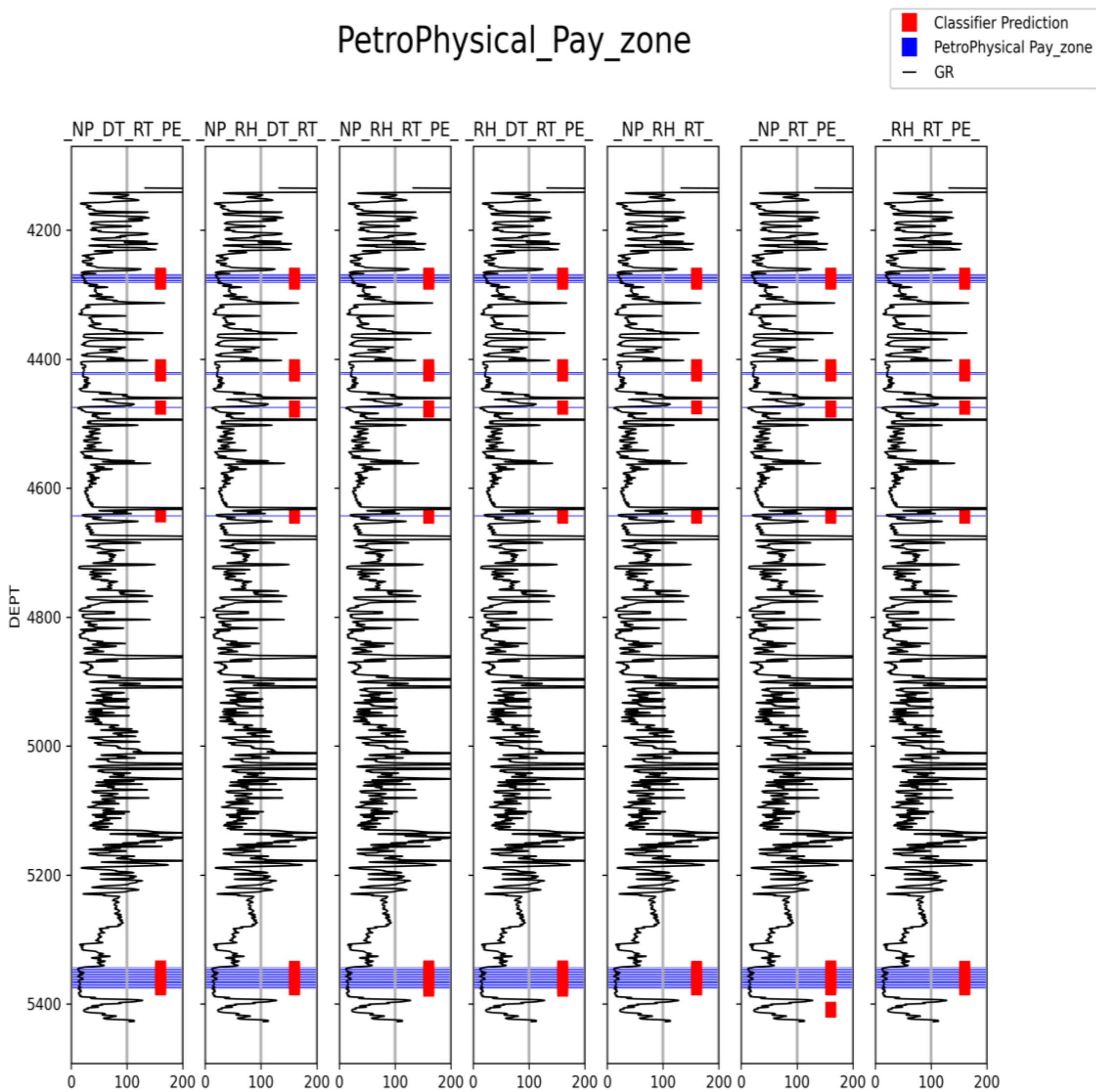

Figure 28 Pay zone prediction Good models basic log included (X, Y GR, DEPT and CALI) part 1

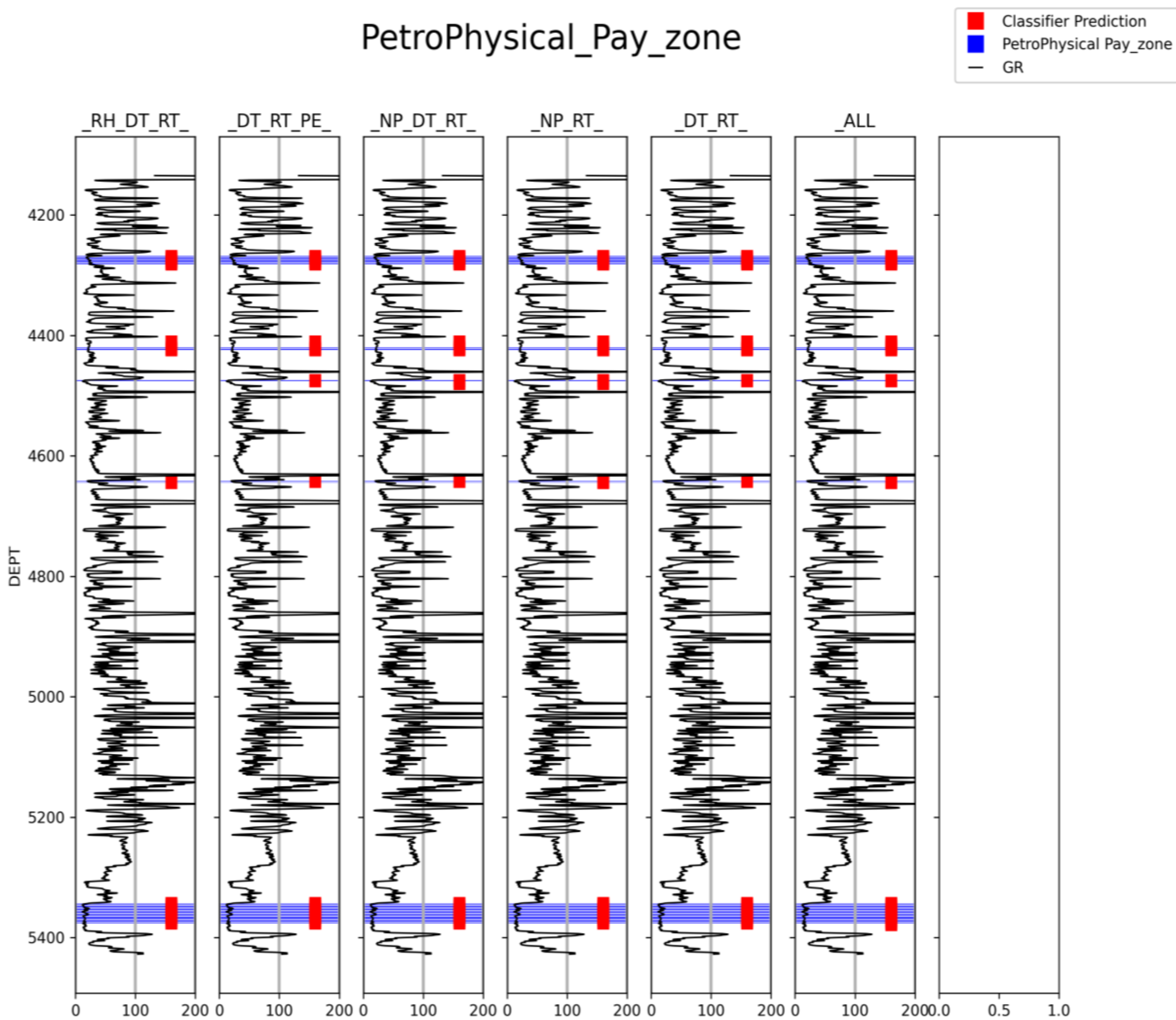

Figure 29 Pay zone prediction Good models basic log included (X, Y GR, DEPT and CALI) part 2

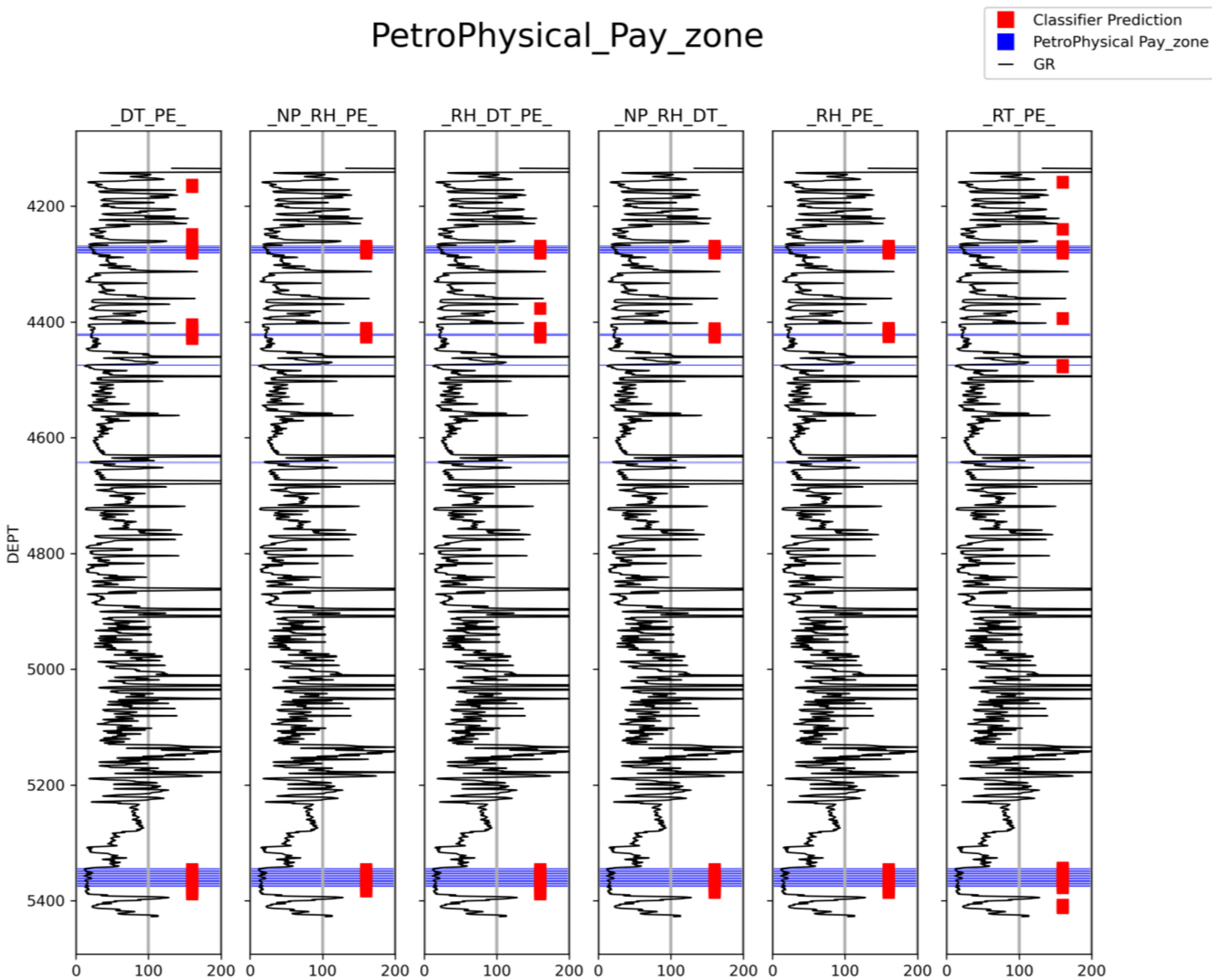

Figure 30 Pay zone prediction Bad models basic log included (X, Y GR, DEPT, and CALI) part 1

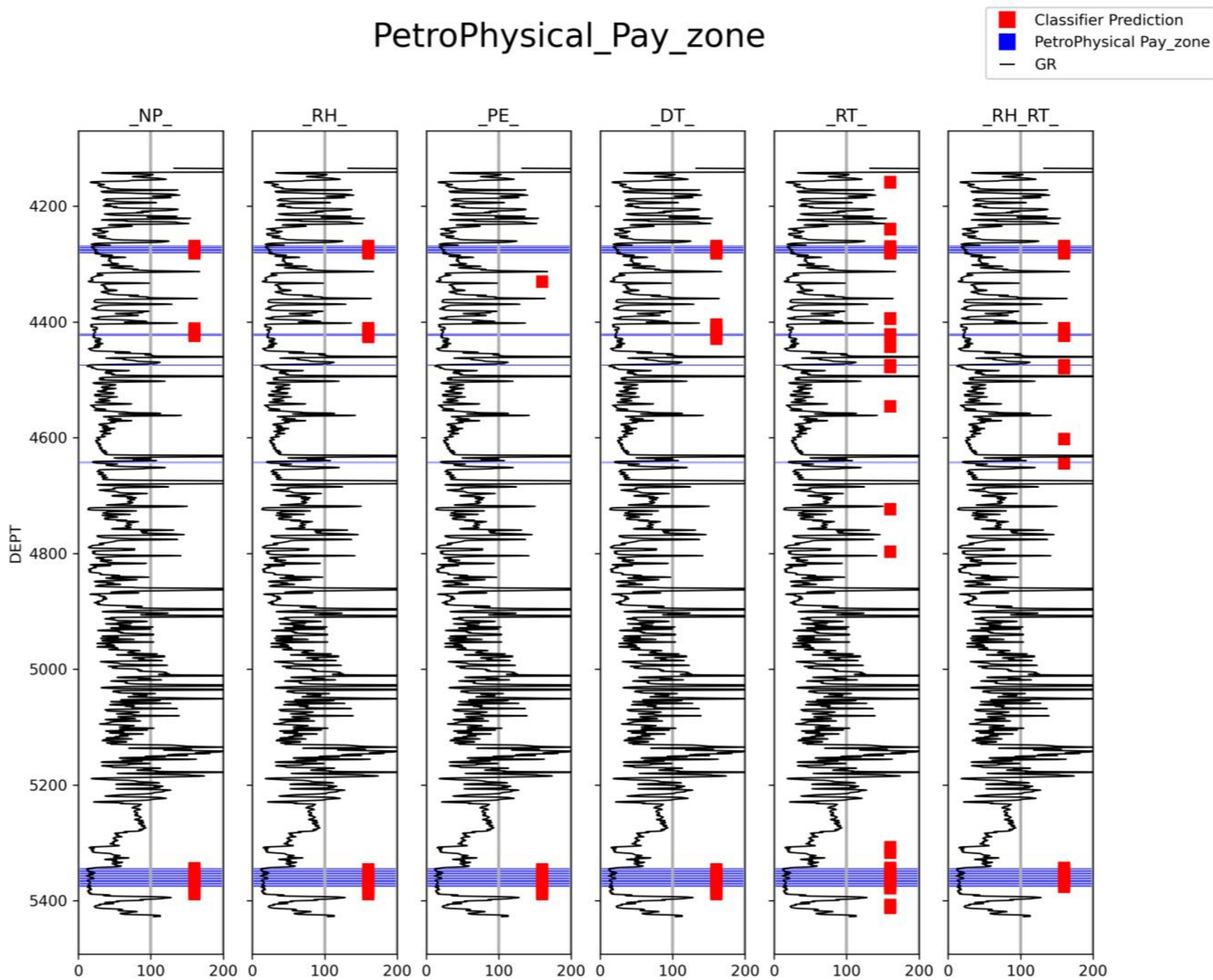

Figure 31 Pay zone prediction Bad models basic log included (X, Y GR, DEPT, and CALI) part 2

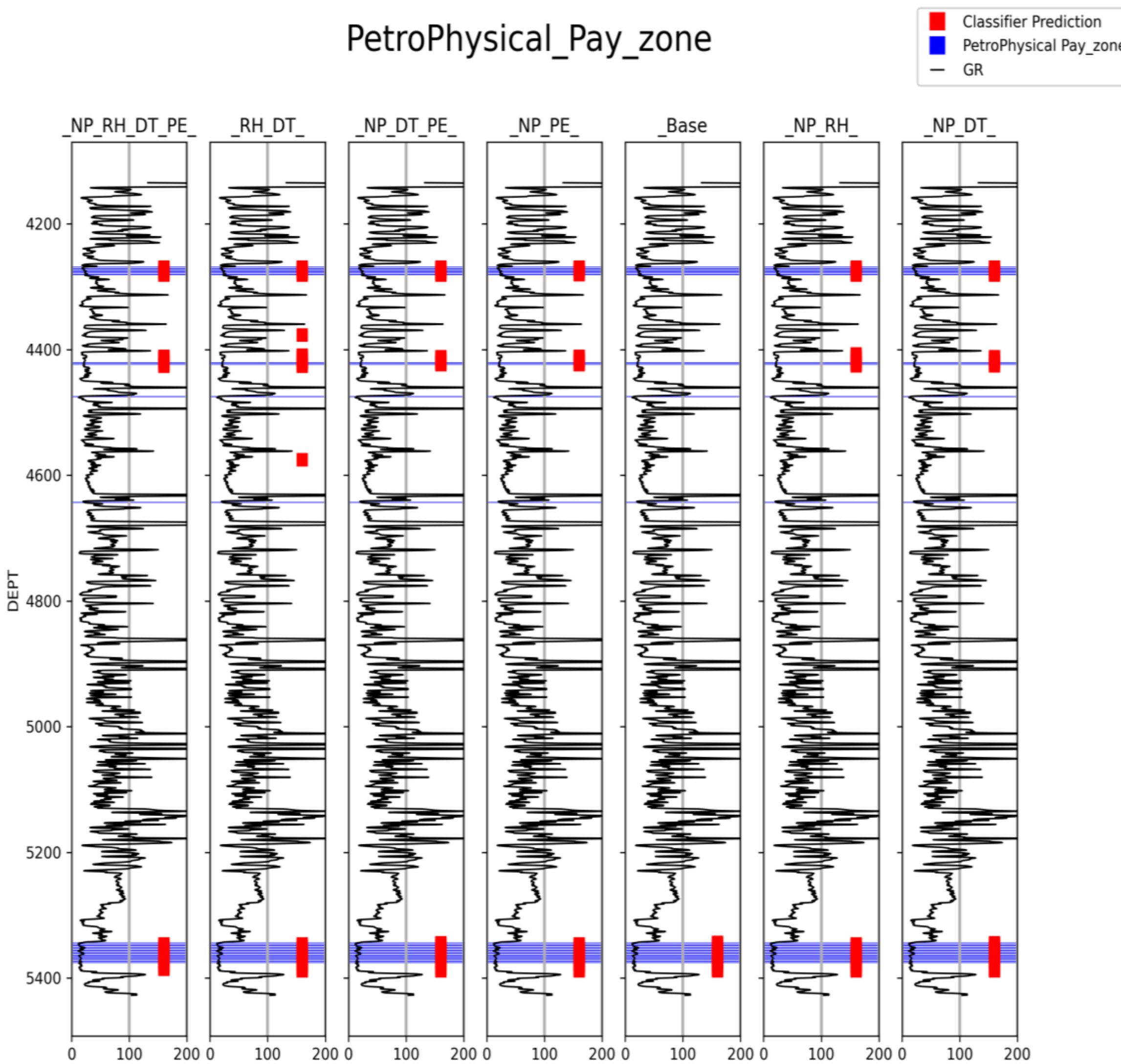

Figure 32 Pay zone prediction Bad models basic log included (X, Y GR, DEPT, and CALI) part 3

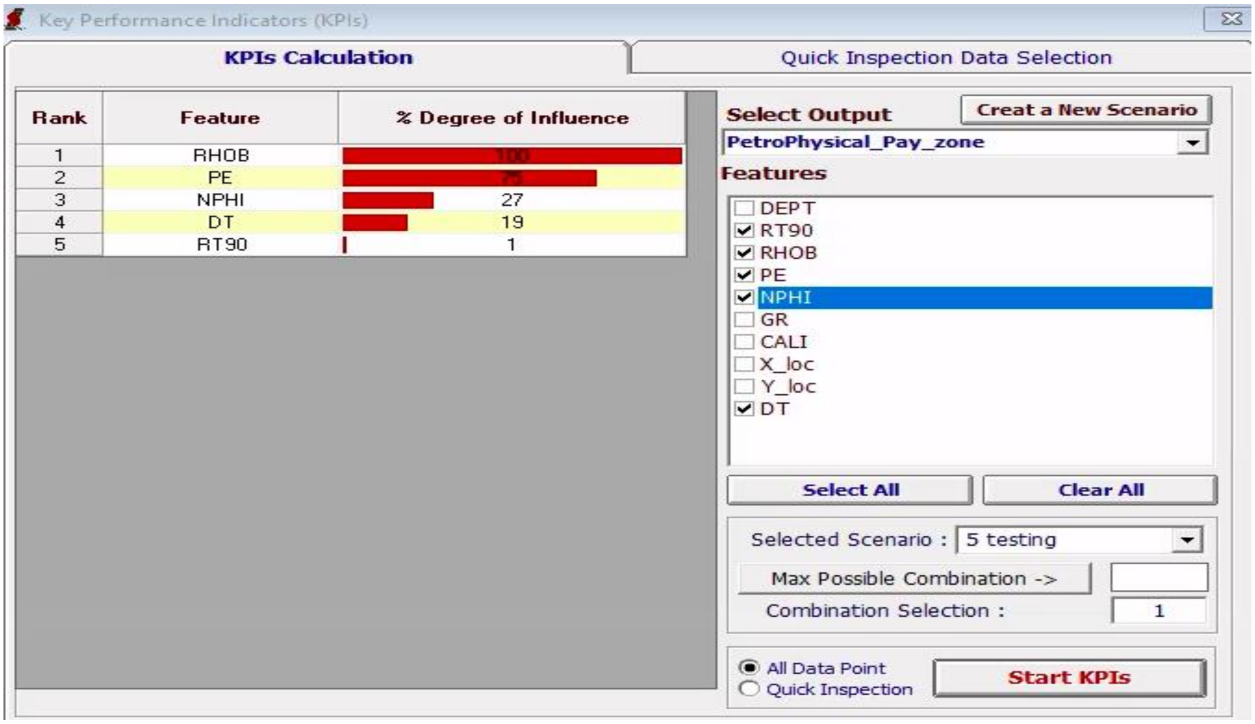

Figure 33 Impact of input log on Petrophysical Pay zone using KPI.

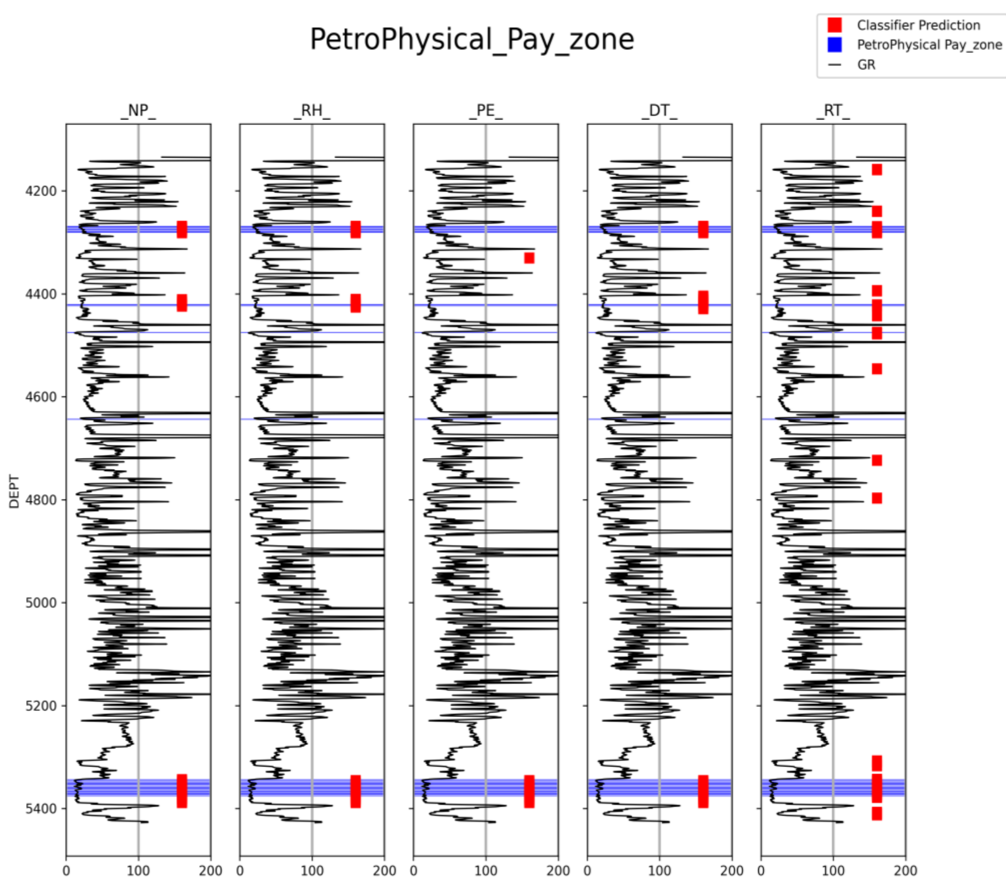

Figure 34 Pay zone prediction single log model basic log included (X, Y GR, DEPT, and CALI)

Discussion:

The base model contains all the logs considered ‘basic’ for our model to function (X, Y GR, DEPT, and CALI), while the ‘ALL’ model includes all the possible logs from our database. Next, the naming of the combination of additional logs was created by using their annotation, such as NP for NPHI or RT for resistivity. A model named _NP_RT contains all the five base logs in addition to NPHI and Resistivity. Using petrophysical analysis, we identified a possible pay zone within our blind well figure 10. Five main pay zone spaces were located as represented by the blue shading on the figures. According to that result, a model considered with good prediction was able to identify these five zones (figures 16 and 17).

Table 3 highlight green models with good prediction. This result identified that all the excellent models required resistivity as an input. Another observation was that all the good models have at least a combination of resistivity and NPHI or resistivity and Delta, except one which is a combination of resistivity, PE, and RHOB, which are the remaining logs to be tested. In addition, a KPI (Key Performance Indicator) was done on the input logs with the petrophysical pay zone as output. Based on figure 15, it was found that RT90 has the most negligible affinity with the pay zone and RHOB the least. This confirmed our model result since KPI only accounts for a one-on-one comparison between an input and an output. In figure 22, we tried to recreate the KPI study with the base log as input and added one of the logs to be tested. It does not lead to a one-on-one comparison with the log of interest and the output as seen with the KPI; however, a similar behavior was observed with the model with only RT added to the base model resulting in a poor prediction and the model with RHOB added giving a better prognosis.

Conclusion

We could closely predict the petrophysical pay zone using AI and machine learning. During this project, a traditional approach to pay zone identification was taken. This required using other software such as PETRA, a minimum of 14 logs used in tracks for visual representation and some petrophysical interpretation. Our result was matched with the pay zone from completion data. Next, the same result was achieved using Artificial

Intelligence with models giving matching predictions with the petrophysical pay zone by only using 7 logs. This shows AI and machine learning ability and accuracy in pay zone identification using fewer resources than traditional methods. A petrophysical background is still required for the data gathering or pre-processing to have an accurate model. However, it might help reduce the expertise gap needed when working in a field with more complex lithology and enable engineers with less experience to predict accurately.

Cited literature

https://glossary.oilfield.slb.com/en/terms/f/formation_evaluation

Archie, G.E., 1942: The electrical resistivity log as an aid in determining some reservoir characteristics.-Trans. AIME, 145, S.54 – 62 (Also in: Trans. Spe, 1941, 146)

Friedman, T. and Smith, M., 2011. the Business Value of Data Quality. (G00218962).

McDonald, A., 2021. When working on petrophysical project a large portion of our time can be taken up getting the data ready for interpretation. URL: https://www.linkedin. com/posts/andymcdonaldgeo_data-datascience- petrophysics-activity-6765699282860306433-xI6W. Posted: [LinkedIn] 2021-2-14.

McDonald, Andrew. "Data Quality Considerations for Petrophysical Machine Learning Models." Paper presented at the SPWLA 62nd Annual Logging Symposium, Virtual Event, May 2021. DOI: https://doi.org/10.30632/SPWLA-2021-0036

Nordloh, Vito Alexander, et al. "Machine Learning for Gas and Oil Exploration." *9th International Conference on Prestigious Applications of Intelligent Systems*, 2020, ecai2020.eu/papers/pais/40_paper.pdf.

Popescu, Marian, Head, Rebecca, Ferriday, Tim, Evans, Kate, Montero, Jose, Zhang, Jiazuo, Jones, Gwynfor, and Geovani Christopher Kaeng. "Using Supervised Machine Learning Algorithms for Automated Lithology Prediction from Wireline Log Data." Paper presented at the SPE Eastern Europe Subsurface Conference, Kyiv, Ukraine, November 2021. DOI: https://doi.org/10.2118/208559-MS

Schon , Jurgen. "Introduction." *Basic Well Logging and Formation Evaluation*, Bookboon.com, 2015, pp. 8–10.