The FORCE 2020 Machine predicted Lithofacies competition results
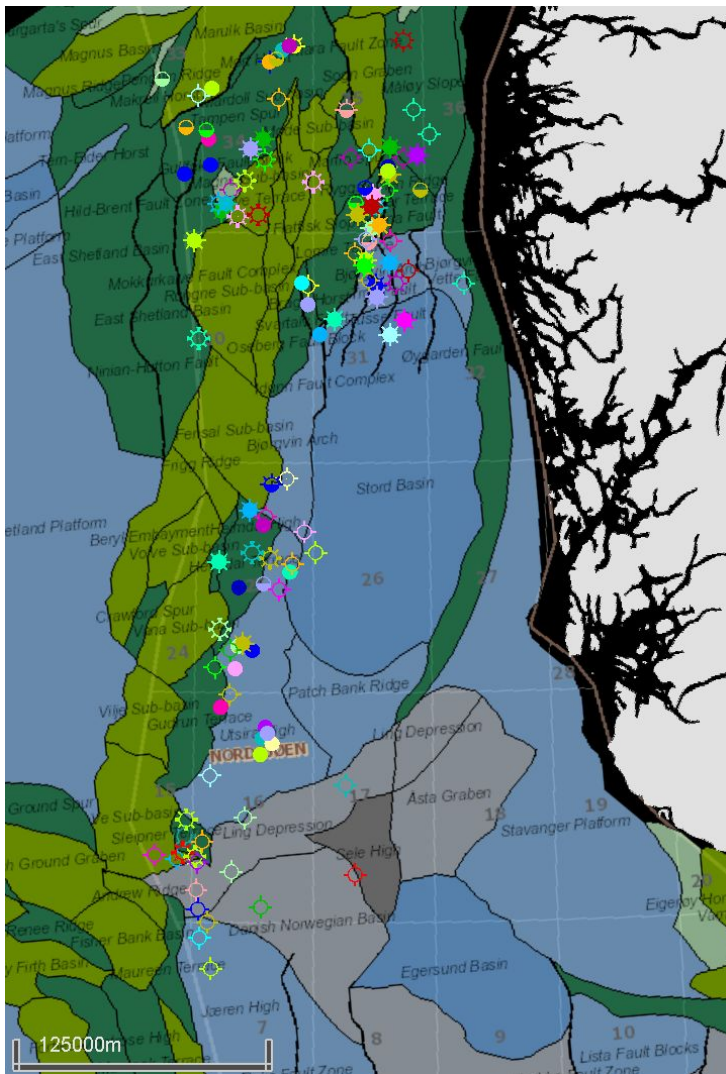
And the winner is….. …the human with the machine

329 teams from around the world signed for the competition. 148 of them submitted at least one solution. 2200 solutions were scored against the blind well dataset of 10 wells. In the end there can be only 1 winner.

1) **Olwale Ibrahim**, an applied geophysics student from the Federal University of Technology Akure, Nigeria won the competition.

2) In second place is the **GIR team** a research team from Universidade Estadual do Norte Fluminense (UENF), located in Macaé, Brazil. The effort was headed by Lucas Aguiar

3) In third place came the **Lab ICA Team** at the Pontifical Catholic University of Rio de Janeiro which was headed by Smith W. A. Canchumuni. Laboratotio de Inteligencia Computacional Aplicada PUC-Rio

Final Score	My team name /personal name is	Score on hidden dataset	My current score XEEK leader board is	My current position XEEK leader board is
1	Olawale Ibrahim	-0.469	-0.5118	24
2	GIR TEAM	-0.4792	-0.5037	11
3	Lab.ICA-Team / Smith A.	-0.49536	-0.4943	6
4	H3G (Haoyuan Zhang, Harry Brandsen, Gregory Barrere, Helena Nandi Formentin)	-0.504489	-0.509	17
5	ISPL Team	-0.50835	-0.4885	2
6	Jiampiers C.	-0.50886	-0.5014	9
7	José Bermúdez	-0.509061	-0.5052	14
8	Bohdan Pavlyshenko	-0.51713	-0.5112	22
9	Jeremy Zhao	-0.51733	0.5264	31
10	Campbell Hutcheson	-0.52206	-0.505	13

As expected the final scores are significantly different on the blind dataset to as compared to the test data. This is a combined effect of the models being overfitted to the data and the blind data not having the same lithology distribution as the combined train and test data (later more on that…). The final scores are incredibly close and when looking at the predictions on the well logs one realises how little difference there is between the top 5-7 models. One can therefore safely say that the top 7 teams are actually all winners.

All data, submitted machine learning codes and final scores are here: https://github.com/bolgebrygg/Force-2020-Machine-Learning-competition

Confusion matrix of the winning model for all wells in the blind /train / test dataset

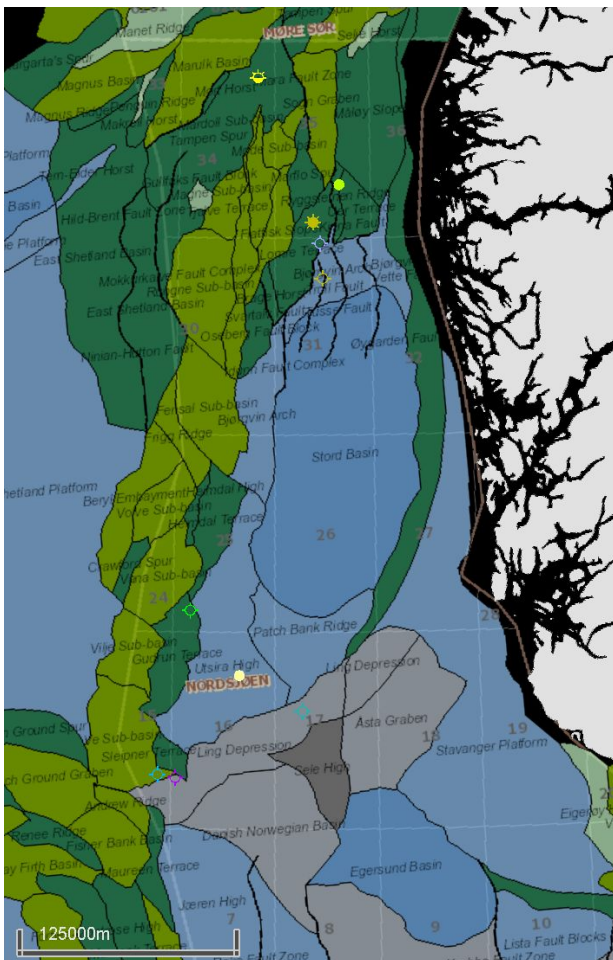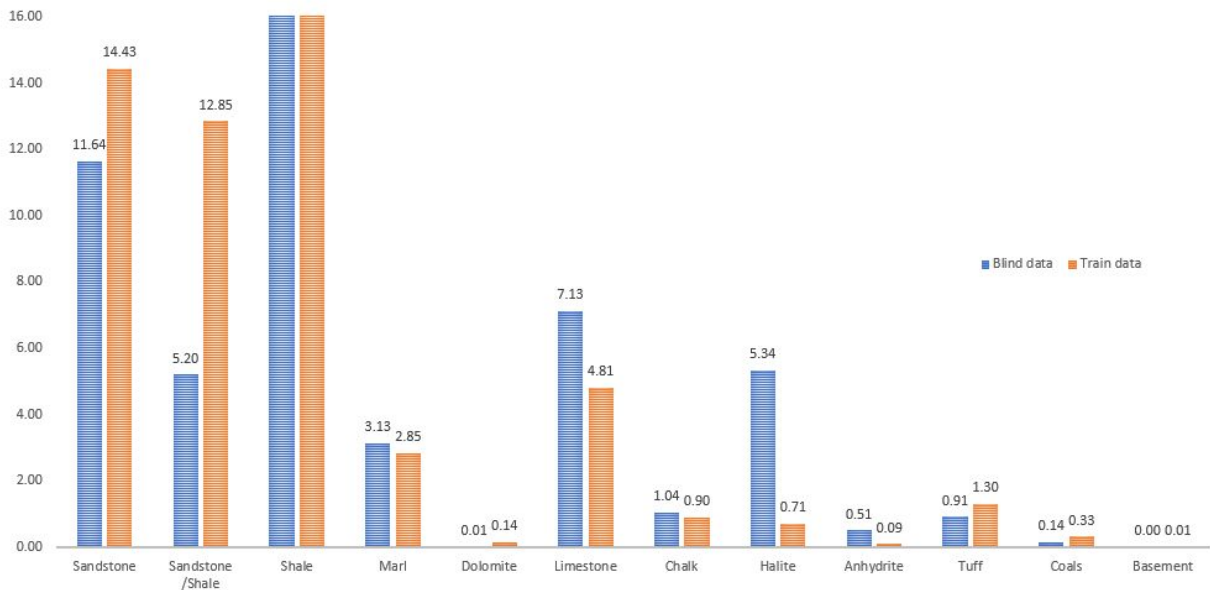| Olawale Predicted	Labeled Sandstone	Labeled Sandstone /S	Labeled Shale	Labeled Marls	Labeled Dolomite	Labeled Limestone	Labeled Chalk	Labeled Halite	Labeled Anhydrite	Labeled Tuff	Labeled Coals	Labeled Basement
Sandstone	86	8	2	1	0	2	0	0	0	2	3	0
Sandstone /Shale	8	80	5	1	0	1	0	0	0	2	4	0
Shale	3	9	91	9	12	4	1	0	1	10	8	0
Marl	0	1	1	79	0	7	3	0	0	0	0	0
Dolomite	0	0	0	0	77	0	0	0	11	0	0	0
Limestone	2	2	1	9	1	80	7	0	2	3	0	0
Chalk	0	0	0	0	0	5	89	0	0	0	0	0
Halite	0	0	0	0	2	0	0	99	1	0	0	0
Anhydrite	0	0	0	0	8	0	0	1	85	0	0	0
Tuff	0	0	0	0	0	0	0	0	0	83	0	0
Coals	0	0	0	0	0	0	0	0	0	0	85	0
Basement	0	0	0	0	0	0	0	0	0	0	0	0

The Dataset

Creating a consistent dataset of this scale is not an easy feat. Not only are there legal issues that need to be considered but the log curves need to be cleaned up and last but not least a consistent interpretation of the lithology needs to be made. We used our funding from the sponsors and FORCE and contracted EXPLOCROWD to make hand crafted lithology interpretation using inhouse data, completion logs, mud logs, and of course the wireline curves. I2G.cloud provided the lithology for 14 wells since they wanted to support the competition. Both companies returned high quality lithology data in a very short amount of time. Congratulations and thanks for that.

The 118 wells dataset spans the South and North Viking graben and penetrates a highly variable geology from the Permian evaporites in the south the the deeply buried Brent delta facies in the North. We held out 10 wells where we only provided the logs (test dataset) and 10 wells that were not provided at all to the contestants. The blind dataset was used to assess the final scores of the supplied models. With hindsight we should have chosen a larger blind dataset since we are not fully representing the lithology distribution of the train dataset with the blind dataset. This introduces a not insignificant element of luck into the final leaderboard. It will be therefore interesting to see how this datasat is being contested, dissected and analyzed in the future. We also hope to augment the dataset with more data. Contact us if you are interested to sponsor (-:

The dataset is generally of high quality, but it is not without faults. It is still based on a semi subjective "interpretation" of various data types . It is also human based and therefore unfortunately error prone. In the analysis section we will highlight some of these shortcomings and outline potential solutions to these.

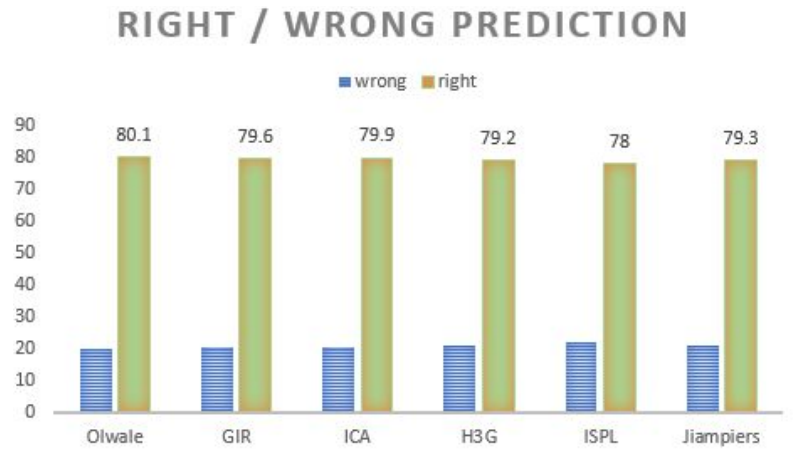PERCENTAGE LITHOFACIES DISTRIBUTIONS TRAIN AND BLIND DATA

An investigation of the provided training data as well as the blind data clearly shows that the lithologic record offshore Norway is dominated by shales and shaly sediments. This is followed by sandstones, limestones, marls and the tufts. The figure below also illustrates that there is an imbalance between the train and the blind dataset where the blind dataset has quite a lot more Halite and Anhydrite while it contains on average less sand and less Sandstone /shale.

The 10 blind wells area reasonably evenly distributed throughout the area of interest .

Human against the Machine- Who wins? -
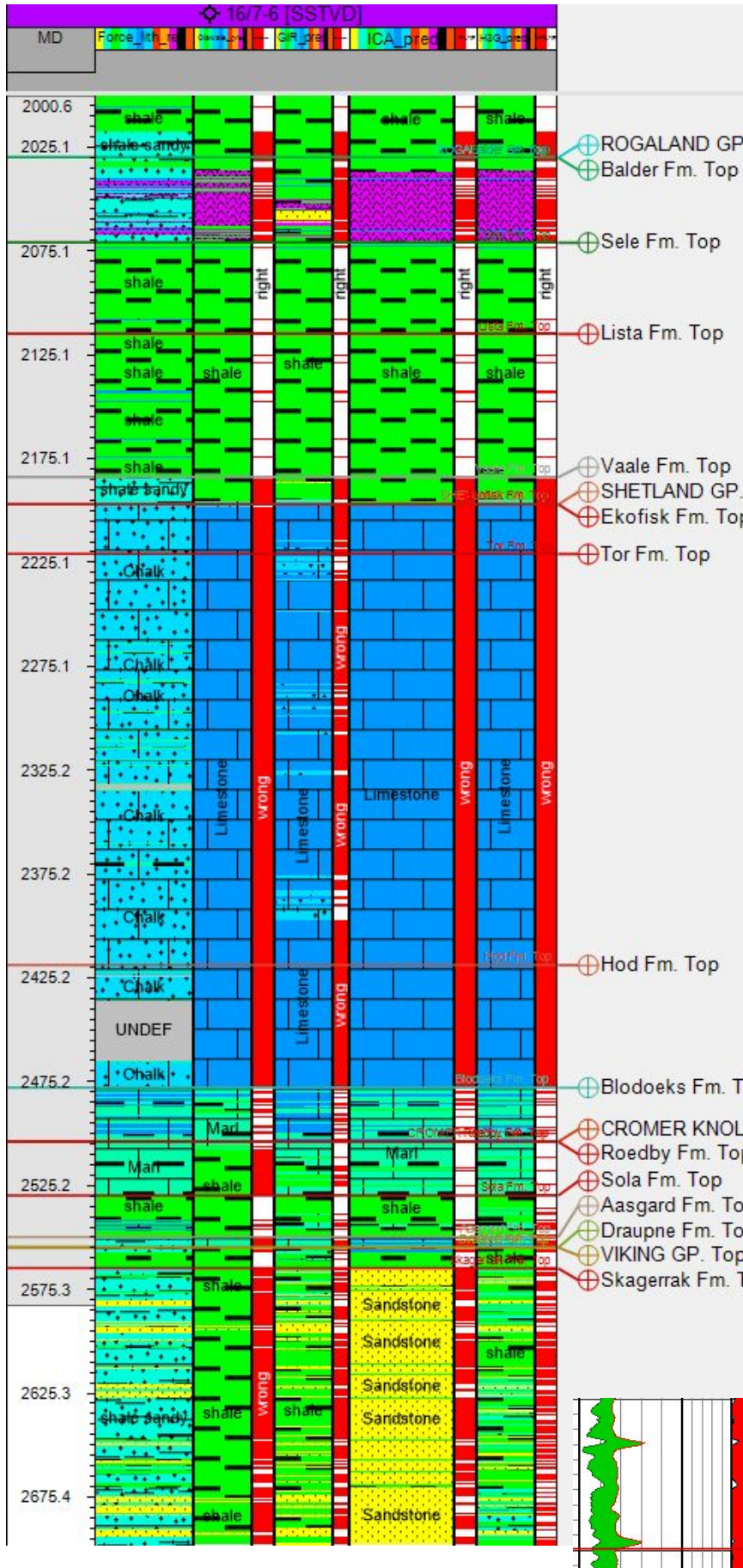# In depth analysis of the lithology predictions by the machine

The winner in this case is clear: The human with the machine!

So how good is the machine at predicting lithology and what can we learn from that?
Even the best model only managed to achieve 80% right predictions on the blind data. The runner up models all scored between 78 and 79 % . The third ranked model even achieved a higher percentage of right predictions than the second place model but was punished a bit harder on getting some obvious lithologies wrong due to the petrophysically based scoring matrix that was used. It also illustrates that the top models are incredibly close in terms of their solutions and only minor details decided in the end who will be the winner.

RIGHT / WRONG PREDICTION

scoring matrix for petrophysical interpretation

label \ prediction	Sandstone	Sandstone/Shale	Shale	Marl	Dolomite	Limestone	Chalk	Halite	Anhydrite	Tuff	Coal	Crystalline Basement
Sandstone	0	2	3.5	3	3.75	3.5	3.5	4	4	2.5	3.875	3.25
Sandstone/Shale	2	0	2.375	2.75	4	3.75	3.75	3.875	4	3	3.75	3
Shale	3.5	2.375	0	2	3.5	3.5	3.75	4	4	2.75	3.25	3
Marl	3	2.75	2	0	2.5	2	2.25	4	4	3.375	3.75	3.25
Dolomite	3.75	4	3.5	2.5	0	2.625	2.875	3.75	3.25	3	4	3.625
Limestone	3.5	3.75	3.5	2	2.625	0	1.375	4	3.75	3.5	4	3.625
Chalk	3.5	3.75	3.75	2.25	2.875	1.375	0	4	3.75	3.125	4	3.75
Halite	4	3.875	4	4	3.75	4	4	0	2.75	3.75	3.75	4
Anhydrite	4	4	4	4	3.25	3.75	3.75	2.75	0	4	4	3.875
Tuff	2.5	3	2.75	3.375	3	3.5	3.125	3.75	4	0	2.5	3.25
Coal	3.875	3.75	3.25	3.75	4	4	4	3.75	4	2.5	0	4
Crystalline Basement	3.25	3	3	3.25	3.625	3.625	3.75	4	3.875	3.25	4	0

So how wrong is wrong? Three wells will be discussed (16/7-6 , 31/2-10, 16/2-7) in detail . The predictions shown are from left to right (Force interpretation, Olawale, GIR, ICA, H3G).
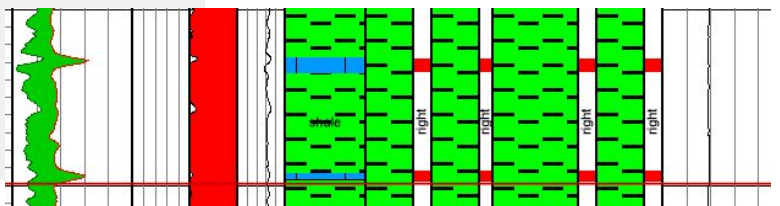
Looking at the amount of red in the right /wrong column in well 16/7-6 one could at first think that machine learning on well data is a useless entertainment for the bored data scientist. Over 60 percent of the predictions in this well are wrong.
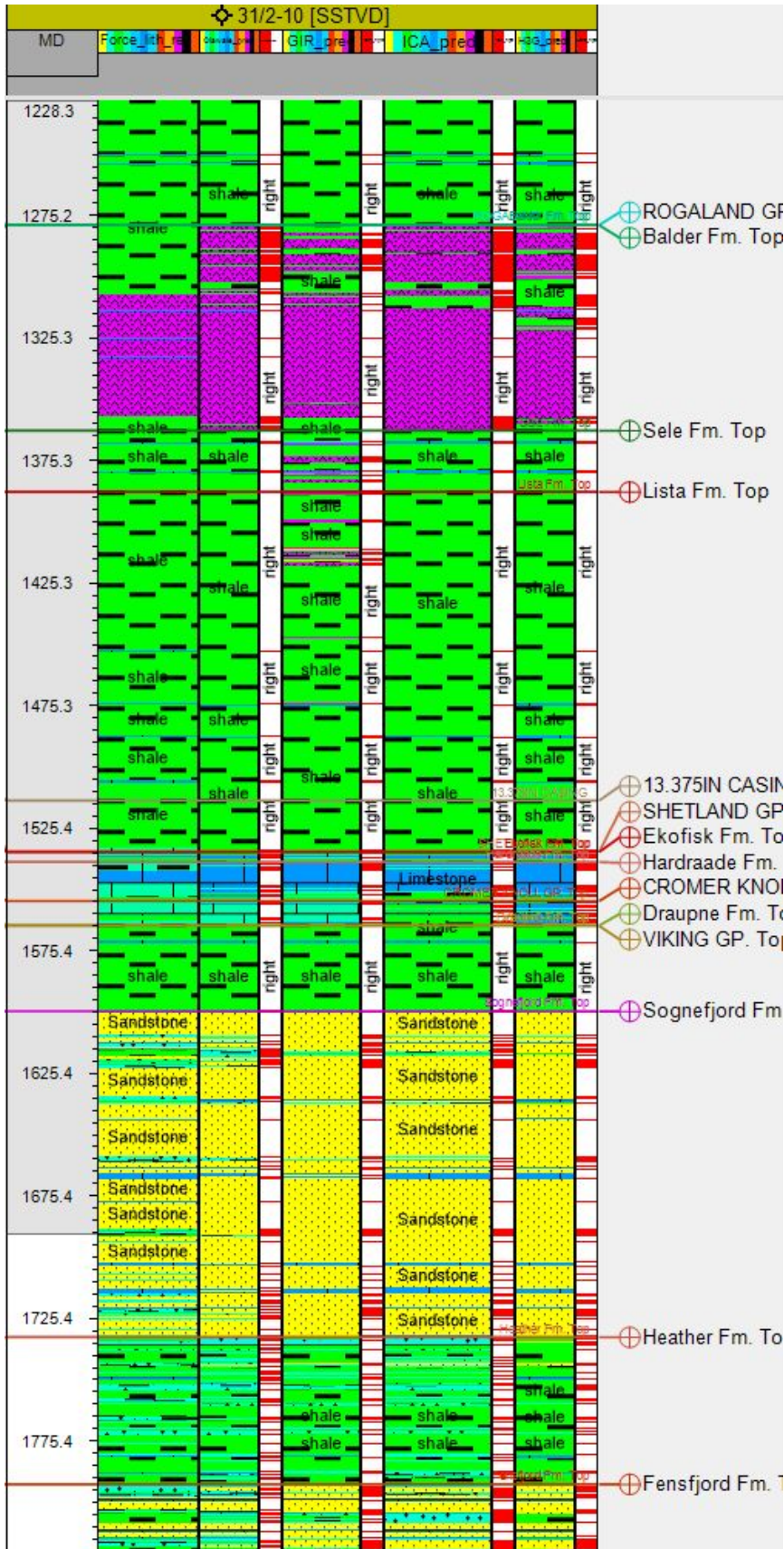
A second look reveals that the predictions might not be that bad at all. Mistaking a limestone with a chalk is entirely permissible.

Where the models really struggle is in the finer details of separating a silty shale from a shaly silt or a shaly sandstone. The models here illustrate that this can be interpreted either way and it is clear that even experts have difficulty to define the exact lithology in these heavily mixed formations like for example the Skagerrak.

The GIR team and the H3G team came up with good predictions.

Interestingly no team built a proper algorithm to map stringers as is shown below.

The 31/2-10 well shows a slightly above average prediction accuracy for the wells in the blind dataset.

One could go as far as saying that the machine makes a 95% percent interpretation and that the minor inconsistencies to the label data need to be checked. It is entirely possible that the label data is wrong.

Where exactly the boundary between a limestone and a marl lies is hard to define on log data alone.

Equally defining the boundaries between a dirty sandstone and a clean sandstone is not easy and often subject to a large degree of individual judgement.

In general the submitted ML models here are more polarizing than the human interpreters

The Rotliegendes intervall in well 16/2-7 (Johan Sverdrup field) illustrates how useful ML can be.

Intriguingly none of the models picked the apparent limestone on the center of a sandstone sequence.

Going back to the mud logs it turns out that the limestone label supplied by FORCE is entirely wrong. The FORCE interpreters were guided by the wrong lithology symbol used on the mudlog

It turns out that this interval is cored and the entire core consists of a lovely conglomerate and breccia as is shown in the image below.

This is just one example where all the machine learning models disagreed with the label given by FORCE. In several of these cases of total disagreement it was found that the label can be disputed and the recently acquired NOROG cuttings images helped to resolve an apparent interpretation conflict.

Conclusions from a regional geologists point of view. Should we use this machine generated data? (personal view of peter bormann)

When analysing the first results I was shocked how poor the machine performed until I realized that I got the legend wrong….. (-:

After the fixing the legend I was intrigued by the result but not entirely convinced. I then started to cross validate some of our labels with the cuttings images and more detailed mud log descriptions.

It appears that the machine provides something like a 80-95% percent solution in most cases. It can be questioned if a 100% solution can ever be achieved given the systemic uncertainty in assigning lithology labels in the first place.

Real ground truth data is hard to come by, with core data being too biased towards sands and the cuttings data suffering from vertical resolution loss.

With this respect it is also not surprising that no team managed to get much above the 80% score on the blind dataset. The 80% may represent the combined uncertainty in the input labels, the log curves and the rocks types that escape absolute definitions. I am however hopeful that this postulation will be challenged in the future.

Comparing the machine predictions to inhouse data and vendor purchased data I came to conclude that the machine predictions offer an extremely valid second opinion and often highlight the very clear shortcoming in these large databases that have been curated, somewhat inconsistently, over the past 50 years.

I think this competition is exemplary on how machine learning can both effectivize and enrich our geoscience workflows. Running all models over a data set and then analysing agreement and disagreement with the original labels or interpreters opinions offers an effective way of ploughing through large datasets while at the same time helping with ideation.

One can analyse for example if the machine predicted missed sand is real and if it turns out to be real start to dream about new exploration opportunities.

Yes, we must (and want to) use this machine generated data but in conjunction with the questioning and knowledgeable geoscientist!

More analysis of the submissions by the winning teams

There are very clear rules on how to determine the winner and those will remain unchanged. The results of the top teams are however incredibly close and therefore it is interesting to investigate in detail why Olawale won and if it really is the best model for general lithology prediction for well logs in the North Sea.

Looking at the confusion matrix (percentage) on the blind data below gives an indication where the models struggle to correctly predict the lithofacies types. All models struggle to correctly identify the Dolomites (1700 samples in train.csv) as well as the Tuffs and the Coals. As discussed before it is apparent that it is difficult for the models to assign the Sandstone /Shale (dirty sands and sandy shales). The reason for this to an extent is the uncertainty of the label itself as well as the algorithms.

Looking just at the confusion matrices the 4th placed H3G model seems to produce the most balanced outcome yet Olwale fared better on this blind dataset because of the differential penalty of the scoring matrix that was introduced.

1st Olawale Predicted

	Labeled Sandstone	Labeled Sandstone /Shale	Labeled Shale	Labeled Marls	Labeled Dolomite	Labeled Limestone	Labeled Chalk	Labeled Halite	Labeled Anhydrite	Labeled Tuff	Labeled Coals	Labeled Basement
Sandstone	78.5	20.5	1.4	5.5		2.7						
Sandstone /Shale	12.7	55.5	8.3	4.4		0.3				11.9	5.7	
Shale	3.6	14.1	86.3	32.9		2.7	2.4	0.2	1.4	21.0	13.8	
Marl	0.1	7.7	2.0	30.7		11.7	5.1					
Dolomite	0.0		0.2	0.0	5.6	0.1		0.0	18.5			
Limestone	4.1	1.9	1.4	26.2	11.1	60.9	15.6		4.5	3.3		
Chalk	0.5		0.0	0.2		21.1	76.9					
Halite			0.0	27.8	0.1			98.9	1.0			
Anhydrite	0.1		0.0	55.6		0.4		0.9	74.6			
Tuff	0.0		0.3			0.1				63.8		
Coals	0.2	0.3	0.1			0.0					80.5	
Basement												

2ndGIR Predicted

	Labeled Sandstone	Labeled Sandstone /Shale	Labeled Shale	Labeled Marls	Labeled Dolomite	Labeled Limestone	Labeled Chalk	Labeled Halite	Labeled Anhydrite	Labeled Tuff	Labeled Coals	Labeled Basement
Sandstone	75.0	24.5	1.5	0.6		2.8					0.8	
Sandstone /Shale	12.8	51.8	6.6	6.6		0.3				5.2	9.2	
Shale	5.0	21.4	87.1	21.7	8.4	2.8	2.6	0.0	1.3	36.2	14.2	
Marl	0.1	1.6	2.2	42.4	1.0	9.5	6.4			0.8		
Dolomite	0.0		0.2	0.1	40.6	0.1			6.0			
Limestone	5.5	0.5	1.4	27.3	6.4	64.2	1.2		8.5	4.6		
Chalk	0.4		1.3			19.7	89.8					
Halite	0.1		0.3	13.4	0.0			98.8	0.2			
Anhydrite	0.1	0.0	0.0	30.2		0.4		1.2	84.0			
Tuff	0.9	0.0	0.6			0.1				53.3		
Coals	0.1	0.2	0.0								75.7	
Basement												

3rd ICA Predicted

	Labeled Sandstone	Labeled Sandstone /Shale	Labeled Shale	Labeled Marls	Labeled Dolomite	Labeled Limestone	Labeled Chalk	Labeled Halite	Labeled Anhydrite	Labeled Tuff	Labeled Coals	Labeled Basement
Sandstone	71.3	19.2	2.4	8.6		7.6			0.6			
Sandstone /Shale	15.1	58.6	8.2	2.9		0.7				11.0	10.4	
Shale	4.7	12.9	86.1	22.1		3.0	0.4	0.2	2.1	23.8	18.9	
Marl	1.0	8.8	1.6	48.0		12.8	2.1					
Dolomite	0.0		0.2	0.2	0.3		0.1		20.9			
Limestone	6.2	0.5	1.3	18.1		58.1	17.1		3.3			
Chalk	1.5		0.0	0.0		15.8	80.3					
Halite			0.0			0.0		98.7	1.7			
Anhydrite	0.1		0.0			1.4		1.0	75.4			
Tuff			0.2			0.2				61.2		
Coals	0.1	0.1	0.0								70.7	
Basement												

4th H3G Predicted

	Labeled Sandstone	Labeled Sandstone /Shale	Labeled Shale	Labeled Marls	Labeled Dolomite	Labeled Limestone	Labeled Chalk	Labeled Halite	Labeled Anhydrite	Labeled Tuff	Labeled Coals	Labeled Basement
Sandstone	79.5	14.4	2.3	2.0		5.1				0.4	0.9	
Sandstone /Shale	11.7	64.4	10.1	6.3		0.8				14.2	7.3	
Shale	3.6	10.3	84.0	43.6		3.1	3.3	0.0	4.0	16.8	14.5	
Marl	0.1	6.5	1.9	31.7		10.6	6.0					
Dolomite	0.1		0.2	0.0	40.0	1.3			2.9			
Limestone	3.6	4.1	1.0	16.2	40.0	60.0	16.6		3.1			
Chalk	1.0		0.0	0.1		15.8	74.0					
Halite	0.0		0.1			0.1		99.2	8.0			
Anhydrite	0.1	0.0	0.0	20.0		2.8		0.7	84.8			
Tuff		0.3	0.3			0.3				65.0		
Coals	0.3	0.0	0.1								78.2	
Basement												

Looking at the absolute score per category (number sample*penalty in scoring matrix) between Olawale and H3G reveals that despite H3G balanced look they are really struggling (red numbers) to correctly identify the limestones and marls while they outperform Olawale in better identifying the sandstone/shale category (green numbers). H3G puts a disproportionately large number of shale samples into the sandstone category.

Olwale /H3G comparison	Labeled Sandstone	Labeled Sandstone /Shale	Labeled Shale	Labeled Marls	Labeled Dolomite	Labeled Limestone	Labeled Chalk	Labeled Halite	Labeled Anhydrite	Labeled Tuff	Labeled Coals	Labeled Basement
Sandstone		1772	-2769	354		-942			-4	-23		
Sandstone /Shale	348		-3886	-338		-210				-36	-15	
Shale	56	1423		-1538		-249	-94	48	-8	171	-13	
Marl	6	828	100			-36	-77					
Dolomite	-30		60			-310		8	351			
Limestone	263	15	935	504			-102		105	18		
Chalk	-196			2		406						
Halite	-24		-92		19	-4			-44			
Anhydrite	-4		-52	-8	29	-893		33				
Tuff	3	-24	-61			-63						
Coals	-43	68	7			8						
Basement												

A comparison of Olawale against the GIR team shows that these two teams are indeed very close. GIR generally does better on the less common lithologies but struggles to precisely assign the Sandstone /Shale category compared to Olawale.

Olwale /GIR comparison	Labeled Sandstone	Labeled Sandstone /Shale	Labeled Shale	Labeled Marls	Labeled Dolomite	Labeled Limestone	Labeled Chalk	Labeled Halite	Labeled Anhydrite	Labeled Tuff	Labeled Coals	Labeled Basement
Sandstone		-2182	-270	564		-11						-8
Sandstone /Shale	136		3458	-220		11				303	-45	
Shale	-553	-2845		882	-60	11	-11	48	12	33	-33	
Marl	-3	930	-210		-5	446	-38			-17		
Dolomite	4		-14	-8		-8		8	286			
Limestone	-602	266	84	-60	-29		252		-45	32		
Chalk	95		4	-97		261						
Halite	-32		-716		-83	20			14			
Anhydrite	-8	-12			-166	-4			-44			
Tuff	-288	-6	-734									
Coals	70	11	59			8						
Basement												

It seems like each of these models has their strength in certain categories and a logic next step would be to combine the achieve an ensemble model with a high predictive power.

Analysing the confusion matrix of Olawle's model ran over all wells in the train, test and blind dataset gives a good indication of the overall performance of the model and the likely "correctness" that can be achieved from such a model in a geoscience production setting.

Additional tweaking of the algorithm like attention to thin beds, better predictions for less common lithologies and geographical clustering could potentially push the outcome into the 85-90% score range which will be comparable or better than normal human performance.

Please also have a look at the comparison of the label versus the predictions at the end of the blog post to decide for yourself if you like the interpretations of the machine or not.

Final confusion matrix for all wells in Olawale's model

Olawale Predicted	Labeled Sandstone	Labeled Sandstone /S	Labeled Shale	Labeled Marls	Labeled Dolomite	Labeled Limestone	Labeled Chalk	Labeled Halite	Labeled Anhydrite	Labeled Tuff	Labeled Coals	Labeled Basement
Sandstone	86	8	2	1	0	2	0	0	0	2	3	0
Sandstone /Shale	8	80	5	1	0	1	0	0	0	2	4	0
Shale	3	9	91	9	12	4	1	0	1	10	8	0
Marl	0	1	1	79	0	7	3	0	0	0	0	0
Dolomite	0	0	0	0	77	0	0	0	11	0	0	0
Limestone	2	2	1	9	1	80	7	0	2	3	0	0
Chalk	0	0	0	0	0	5	89	0	0	0	0	0
Halite	0	0	0	0	2	0	0	99	1	0	0	0
Anhydrite	0	0	0	0	8	0	0	1	85	0	0	0
Tuff	0	0	0	0	0	0	0	0	0	83	0	0
Coals	0	0	0	0	0	0	0	0	0	0	85	0
Basement	0	0	0	0	0	0	0	0	0	0	0	0

ASome notes from the teams:

We asked the top 10 teams to write a short paragraph about themselves and the models

Olawale Ibrahim

I am a fifth (final) year undergraduate student of the Federal University of Technology Akure, Nigeria. I have my final year undergraduate project ongoing which is on the integration of deep learning for reservoir characterizations and better formation evaluation.

A 10-fold xgboost stratified cross validation technique was used as the final model. Extensive local validations were done to prevent overfitting the open test LB. 10 random wells from the train data set were used in preparing a validation set. Two validation sets were made from each train set prepared.

Special thanks to the organizers and sponsors for the competition. The competition and data is a great step into ensuring more open source contributions in geoscience both from individuals and O&G companies especially. It was fun participating and I hope to get more opportunities to put the experience gained in solving similar challenges in the future.

Email: ibrahim.olawale13@gmail.com

LinkedIn Profile: https://www.linkedin.com/in/olawale-ibrahim-a3a675175/

GIR team

GIR is a research group from Universidade Estadual do Norte Fluminense (UENF), located in Macaé, Brazil. We work mainly to improve reservoir characterization and management by solving problems related to the integrated analysis of geological, geophysical, and reservoir engineering data.

Our primary concern in the competition was building a robust and efficient classifier to handle the training dataset. We already knew XGBoost fits such a task in connection to standards steps of preprocessing, data imputation, and feature augmentation. Our effort concentrated on training, cross-validation, and testing strategies to fine-tune the classifier. Finally, we adopted a petrophysical perspective to specialize preprocessing to perform feature selection and engineering using wavelet transform to help differentiate specific lithologies.

Lucas Aguiar (lucasaguiar26@gmail.com) / **LinkedIn**:
https://www.linkedin.com/in/lucasaguiar26/

Maurício Matos (mauricio.lenep.uenf@gmail.com). **LinkedIn**:
https://www.linkedin.com/in/engmauriciomatos/

Website: giruenf.org / **LinkedIn**: https://www.linkedin.com/in/giruenf/

ICA team

Smith W. Arauco Canchumuni received the M.Sc. and Ph.D. degree in Mechanical Engineering (2013) and Electrical Engineering (2017), respectively from Pontifical Catholic University of Rio de Janeiro (PUC-RIO), Brazil. He is a Mechatronics Engineer who graduated from the National University of Engineering - UNI-Peru (2009).

Currently, working as a researcher at Applied Computational Intelligence Laboratory (ICA).

For the purpose of filling the missing values, we use a simple methodology to complete the values with a median or mode, depending on the type of variable (continuous or discrete).

To train, the model only used variables with more than 50% of the data. Also to introduce temporal information was created new features based on the differential value with respect to depth. The training process consists of using the ensemble network, through the scikit-learn library. At the output network, we apply a median filter using a local window-size given by kernel size, to replace the isolate prediction values.

Contact: saraucoc@uni.pe

H3G Team (Equinor)

We are Team H3G (**H**arry Brandsen, **H**aoyuan Zhang, **H**elena Nandi Formentin, **G**regory Barrere) from Equinor. Harry holds a degree in geology but worked most of his career as a petrophysicist and always had a keen interest in coding and digitalization. Haoyuan is a data scientist with a PhD in Bayesian statistics and inference and has worked previously on and is currently involved in other machine learning projects. Helena is a software developer with PhD in statistics and petroleum engineering, making sure not a data point was neglected. And Gregory is a Geo/Data scientist with an MSc in Petroleum Geosciences, on the bridge between geoscience and data analytics/machine learning.

Whilst we are all from Equinor, we had never met each other – nor live or virtually: we're in 4 different departments in 4 different Norwegian cities (Trondheim, Stavanger, Bergen, Oslo). Despite this, collaboration was energetic, smooth and at full speed from the start due to a good combination of vibrant vibes, eager enthusiasm and nitty-gritty knowledge. The team was put together by our internal sponsor, aiming for precisely this: to combine and get the best of both subject matter expert skills as well as pure data science/machine learning knowledge. The wide variety of backgrounds facilitated tremendously performing these loops in an efficient manner

howardhyzhang@gmail.com

ISPL team

The ISPL_Team is made up of researchers and PhD students from the Image and Sound Processing Laboratory (ISPL) of the Politecnico di Milano (Italy). ISPL research focuses on advanced multimedia signal processing and geophysical data processing solutions.

Our approach to the challenge is based on boosted trees; we train different models depending on the features present in the current test well. The final model is obtained by soft voting among the best models in validation

Maykol Jiampiers Campos Trinidad

Mechatronics Engineer at National University of Engineering, Lima, Peru. Student of BI-Master MBA at PUC-Rio from May of this year.

About my solution, for handling null values, I filled them with the median of each numeric feature, and the most frequent value for categorical variables (mode) with one-hot encoding. Also, created some new features such as Medium Porosity (PHIA), Total Organic Carbon (TOC), etc; and took just the most important attributes. In order to train, I choose the default model Random Forest with 5 stratified K-folds (cross validation) and make a kind of ensemble for inference, where the class with the best mean probability was taken. For post-processing, I make a function to avoid isolated labels in windows of 20 measures.

mcampos@uni.pe

José David Bermúdez Castro

José David Bermúdez Castro, Ph.D. and M.Sc. in Electrical Engineering from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio) in 2019, and 2015, respectively. In 2009, he graduated in Electronic Engineering from Universidad Del Norte in Barranquilla. He currently works as a researcher in the Computational Intelligence Laboratory- ICA at PUC-Rio. His research interests include Deep Learning, Pattern Recognition, Computer Vision, and Remote Sensing.

The procedure followed in this work consisted of first filling the missing values of the most relevant features and then training a committee of random forest classifiers. Some of the features were filled by its median statistics, others by replicating the last valid value, up to bottom, or vice versa (edge padding), by first sorting them by the depth and well. Finally, others were estimated using regression models.

Email: bermudezjosedavid@gmail.com

Jeremy Zhao

Jeremy Zhao is a process engineer based out of Calgary, Canada (but currently stuck in Australia after his one year trip around the world was interrupted in March by COVID). This is his first time participating in a machine learning challenge after just participating in a datathon hosted by the Society of Petroleum Engineers in Calgary. This challenge was a steep learning curve because I had very little exposure to machine learning prior, so bouncing ideas off of my previous datathon team helped immensely. An even greater obstacle is the fact that I know very little about lithologies, so luckily there were people that I could go to when it came to the domain expertise of the subject matter. As you can see in my code, there were a lot of aspects that I had to understand rather quickly in order to build a good enough model to predict the lithologies. Learning how to look at data distribution on a logarithmic scale, classifying the formations using a label encoder, knowing how to treat outliers with the help of my colleagues, and imputing (although leaving nulls or blanks as zero was better than imputing) really helped in my understanding of how the machine learning responded. One thing I did not show in my final code was feature engineering and feature importance runs in order to see which variables might help my runs, which is why I ended up including rate of penetration (ROP), because ROP was considered high enough in my opinion to influence the results.

I want to let people know that it is possible to learn machine learning on your own without too much formal education as long as you're willing to put in the time and effort. Knowing who to go to when you're stuck and/or need help is absolutely crucial as well, as I wouldn't have placed top 10 if I didn't have that support system as well.

For contact information purposes, people can reach me on my professional email at jeremy@larcheng.com

Campbell Hutchinson

Description: I work as CCO of a large publicly-traded technology company. I've always been fascinated by the oil and gas industry and hope to eventually go back to university to study engineering. I used a fairly traditional machine learning stack after trying a large number of different options (PyTorch, Fastai, RapidsAI, LightGBM, CatBoost, XGBoost, SkLearn, etc...). In the end, the combination of XGBoost with hyperparameter optimization seemed to be an effective approach that allowed for fast experimentation because XGBoost has GPU capabilities and was effective because tree-based methods seemed effective on the dataset. I also looked at online videos on rock lithology to try to get an idea for what composite features might be useful to add to the model.

I think it would be worth further exploring how to handle datasets like the one in the competition, where it looks like one has a lot of data (1m+ rows) but it is actually not not as much data as it might appear to be (~100 wells) as the wells have a lot of internal similarity and one needs to organize one's validation methods using the well labels. It creates an interesting challenge because it is expensive to do k-fold validation for model hyperparameters with a large number of folds but doing another validation method either risks over-fitting the hyperparameters to the selected validation set or, if one does k-fold validation with a small number of folds, not being representative of the final problem because the fold training data would be much smaller than the final training set (e.g. for 3-fold: ~66 wells fold training, ~100 wells final training).

Bohdan Pavlyshenko,

Data Scientist, PhD, Ukraine, e-mail: b.pavlyshenko@gmail.com

How did we organize this competition?

2020 being under the sign of Corona proved to be a fruitful year to engage into a virtual global competition. We never met once in person and actually never met in person before we started to organise this event.

The fact that everything was virtual actually helped since we could easily schedule meetings between 9 and 11 pm at night when kids are in bed. We had good fun making this happen and faced some hard challenges like a last minute need to create our own lithology label data, getting sponsor money in a year when oil prices turned negative for a while. The legal aspect of using a US based and Shell backed scoring platform (Xeek) turned out to be a very hard nut to crack and we are indebted to Matt Hall from Agile who in the end managed to close the legal deal.

Special thanks go to Gustavo Lopes from Explocrowd who expertly managed the creation of the label dataset and the train, blind, test split.

We are glad and relieved that so many people found this challenge interesting and that we provided the community with a good real world dataset that likely will be used in many universities and companies for training purposes.

It is hard to precisely put an amount of time that we spent organising this event in addition to the day to day work tasks that we have. A rough estimate could be between 20 and 50 working days combined between the five people in the organizing team

All the best for the future

Your FORCE team and supporters

Peder Dischington (NPD)

Surrender Manral (Schlumberger)

Petter Aursand (AkerBP)

Fahad Dilib (Equinor)

Peter Bormann (ConocoPhillips)

Examples of the wells versus the prediction (Force Label on the Left /Olawale prediction on the right) (blind/test train wells)