Generating Value from Publicly Available Oil and Gas Data

Khushal Adlakha, Data Scientist, Pro Well Plan AS

ProWellPlan

Agenda

- Introduction
- Use Case
- Process and Algorithm Used
- Results
- Conclusion

Introduction

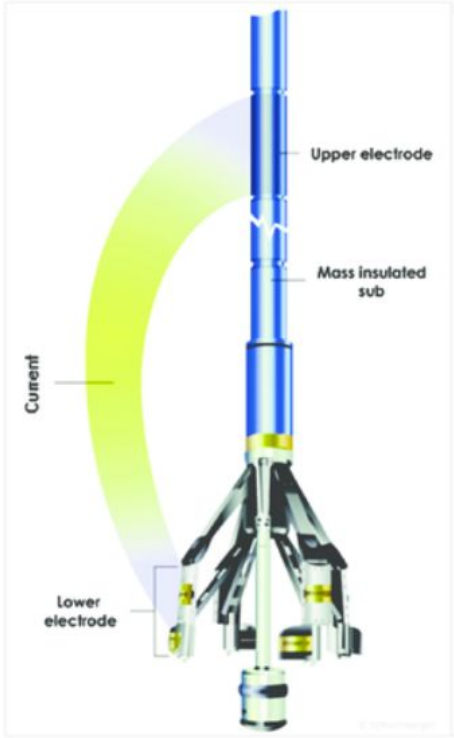
- Study - Mikkel, Petter (NHH) and Pro Well Plan (https://github.com/pro-well-plan/thesis_stratigraphy_prediction_2019)

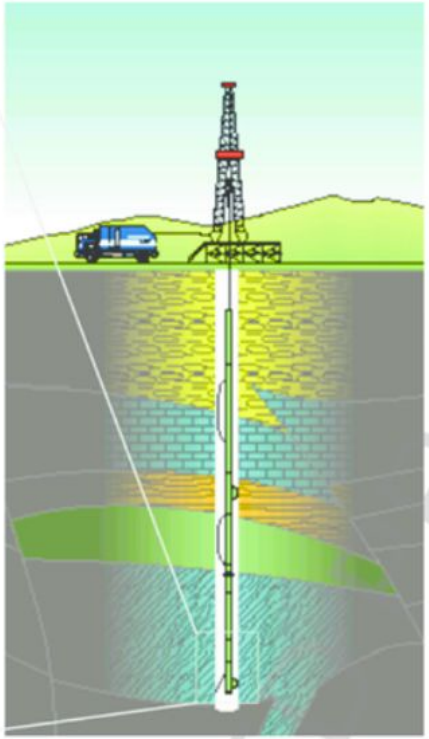
Introduction

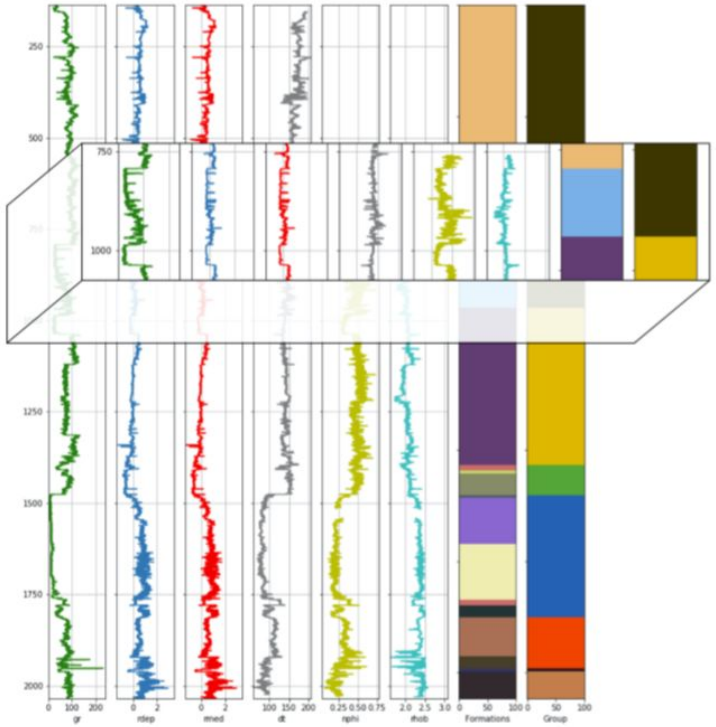
- 9000 wellbores fetched from Norwegian Petroleum Directorate
- Semi-public data

Introduction - Composite Logs

- Tools used for gathering downhole data
- Measures different properties - used for interpretation of rocks

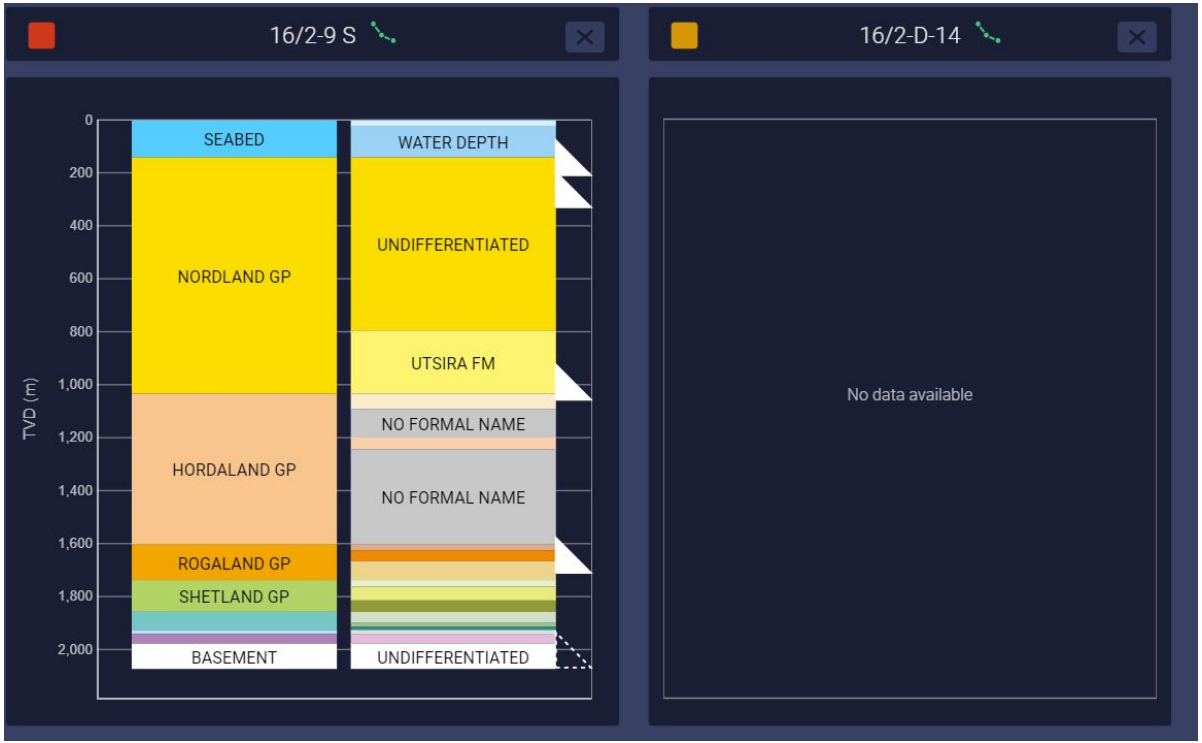




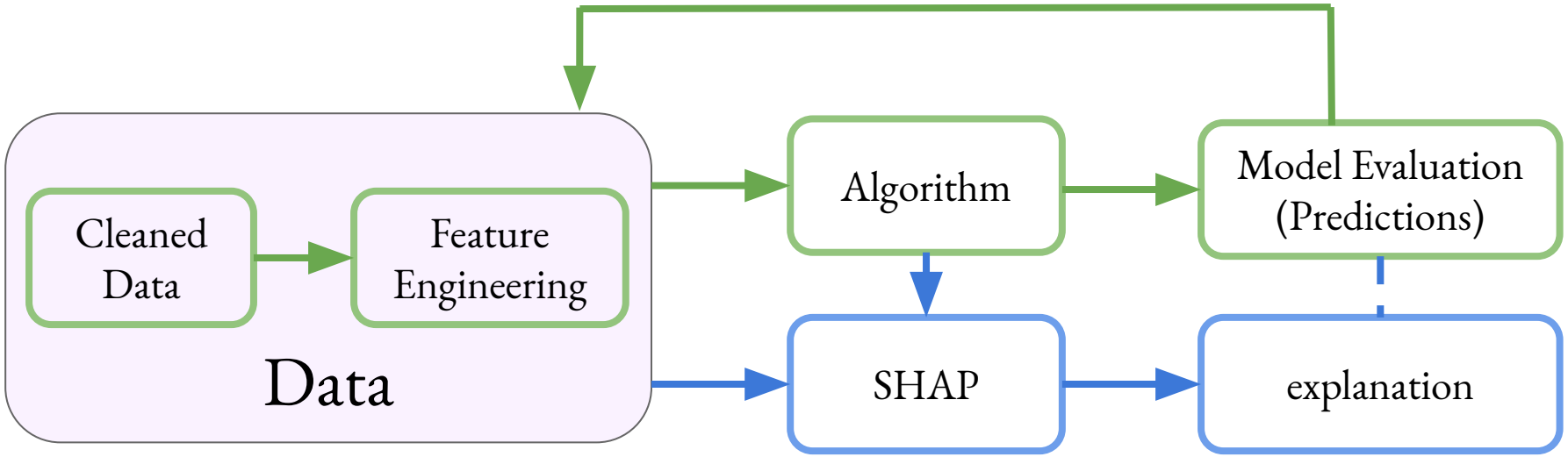


Use Case

- Not all wellbores have predicted formations and groups
- Bu many have composite logs!



Workflow


Cleaned Data
Feature Engineering
Data
Algorithm
Model Evaluation (Predictions)
SHAP
explanation

Algorithm

Logistic Regression

- Base Model - used as **Benchmark Model**
- **Stratified Shuffle Split** for splitting training, validation and test set
- **No** Bayesian Optimization
- **Multi-class** Classification
- **F1** score used to evaluate model

LightGBM 1

- **Hyperapramter Tuning** - Bayesian Optimization
- **Custom Stratified Split**
- **Multi-class** Classification
- **F1** score used for evaluation

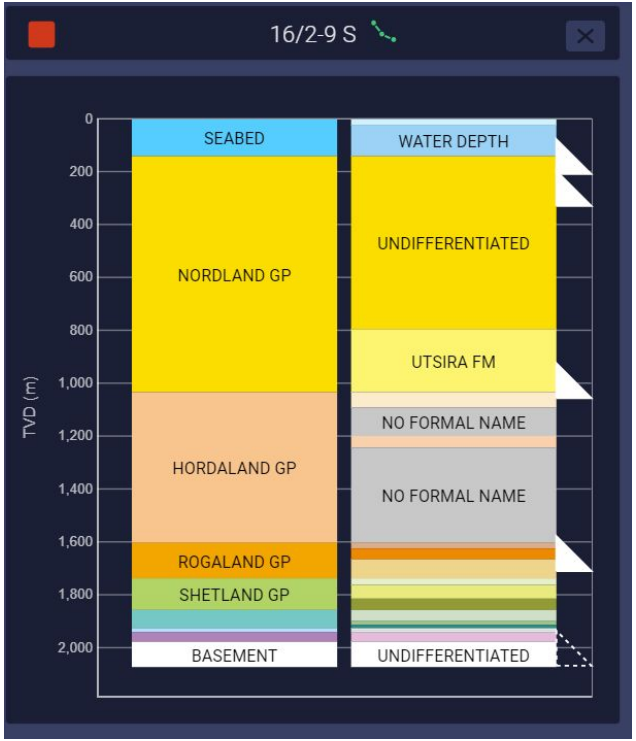
Source

Target Variable

note_id	subject_id	category	note
1	1	ECG	blah ...
2	1	Discharge	blah ...
3	1	Nursing	blah ...
4	2	Nursing	blah ...
5	2	Nursing	blah ...
6	3	ECG	blah ...

https://stackoverflow.com/questions/56872664/complex-dataset-split-stratifiedgroupshufflesplit

Source


16/2-9 S
TVD (m)
0
200
400
600
800
1,000
1,200
1,400
1,600
1,800
2,000
SEABED
NORDLAND GP
HORDALAND GP
ROGALAND GP
SHETLAND GP
BASEMENT
WATER DEPTH
UNDIFFERENTIATED
UTSIRA FM
NO FORMAL NAME
NO FORMAL NAME
UNDIFFERENTIATED

Algorithm

Logistic Regression

- Base Model - used as **Benchmark Model**
- **Stratified Shuffle Split** for splitting training, validation and test set
- **No** Bayesian Optimization
- **Multi-class** Classification
- **F1** score used to evaluate model

LightGBM 1

- **Hyperapramter Tuning** - Bayesian Optimization
- **Custom Stratified Split**
- **Multi-class** Classification
- **F1** score used for evaluation

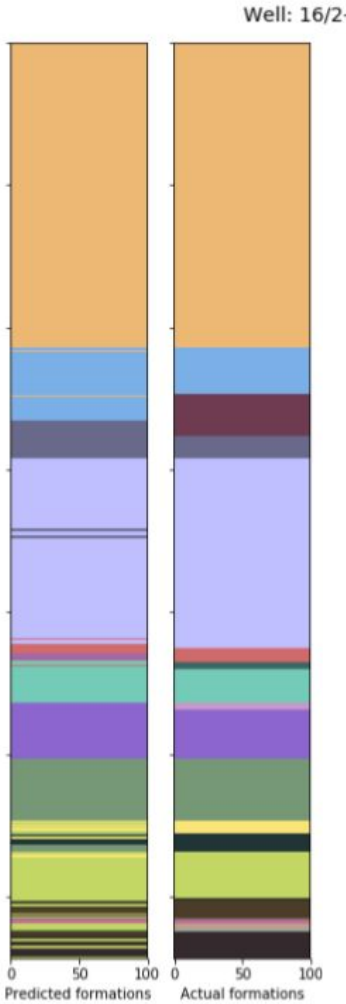
LightGBM 2

- **Stratified Shuffle Split** used instead of custom function

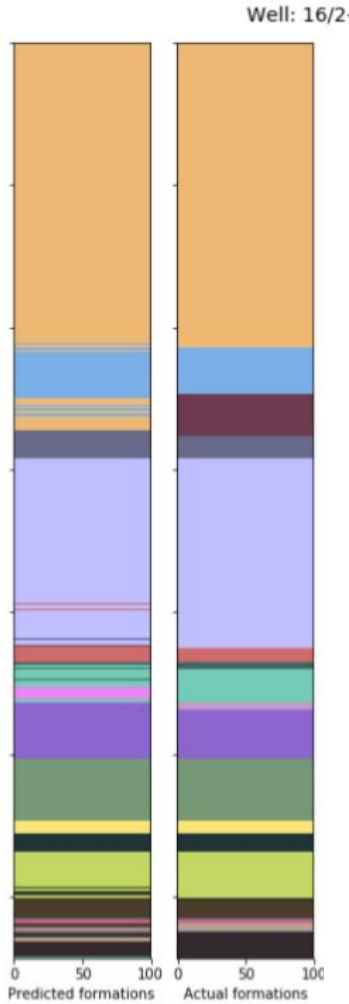
LSTM

- Formations close to each other are likely to depend on each other and hence memory network (**depth vise sequential information**)
- Not much time spent on this model

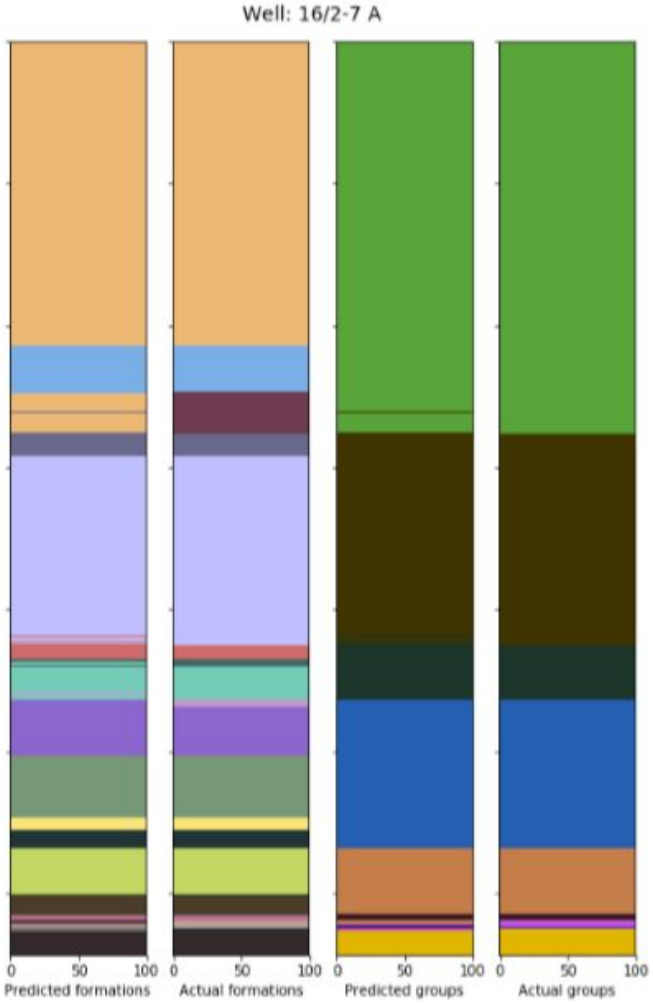
RESULTS



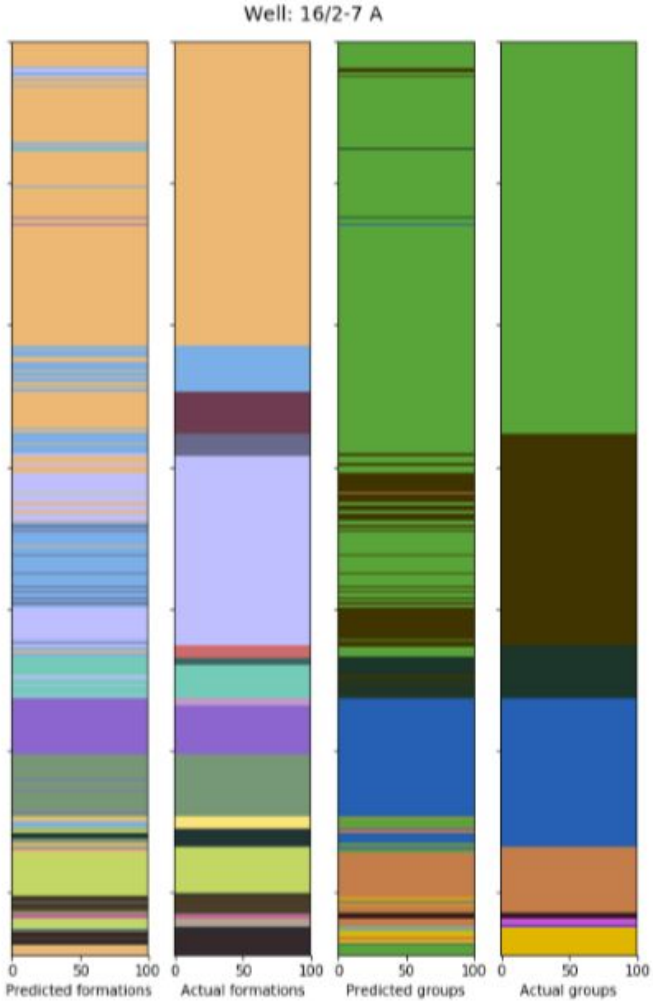
(a) Logistic Regression



(b) LightGBM 1



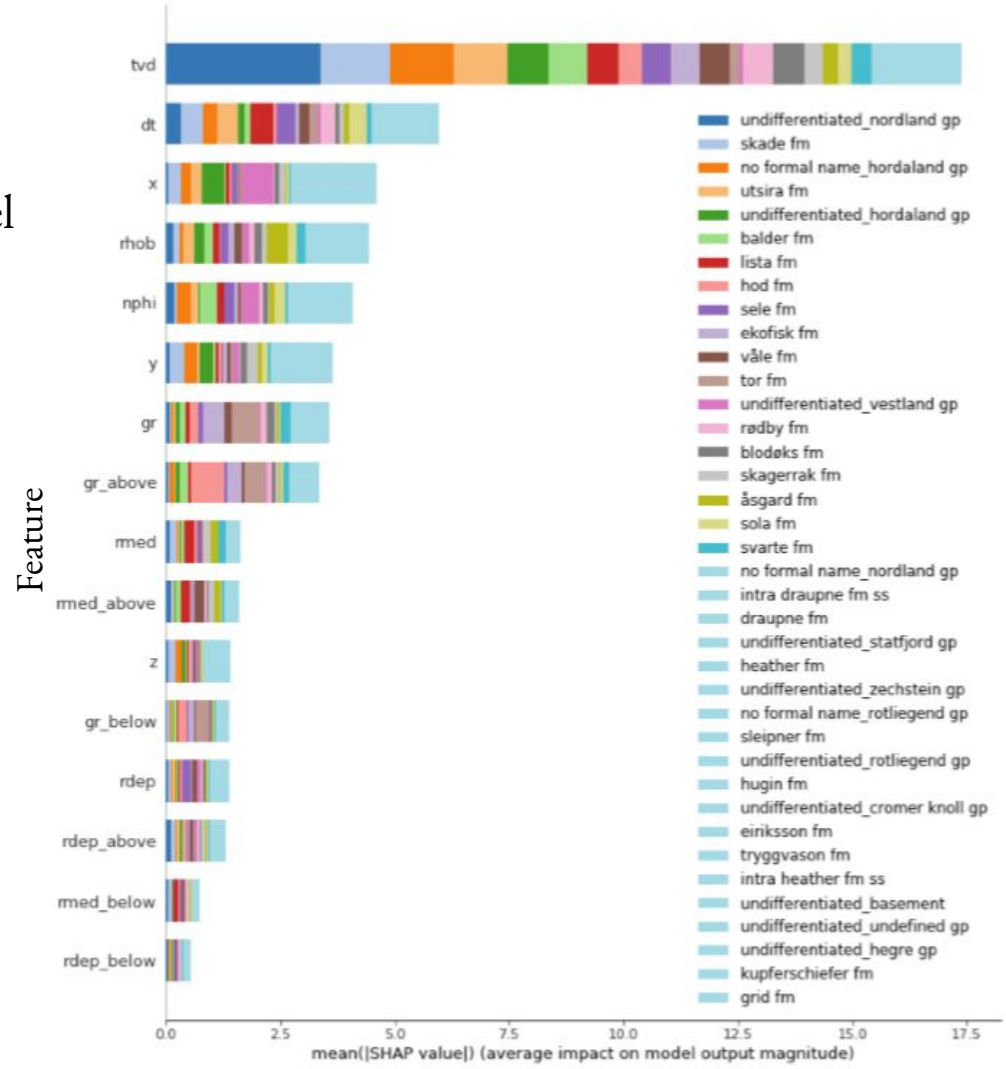
(c) LightGBM 2



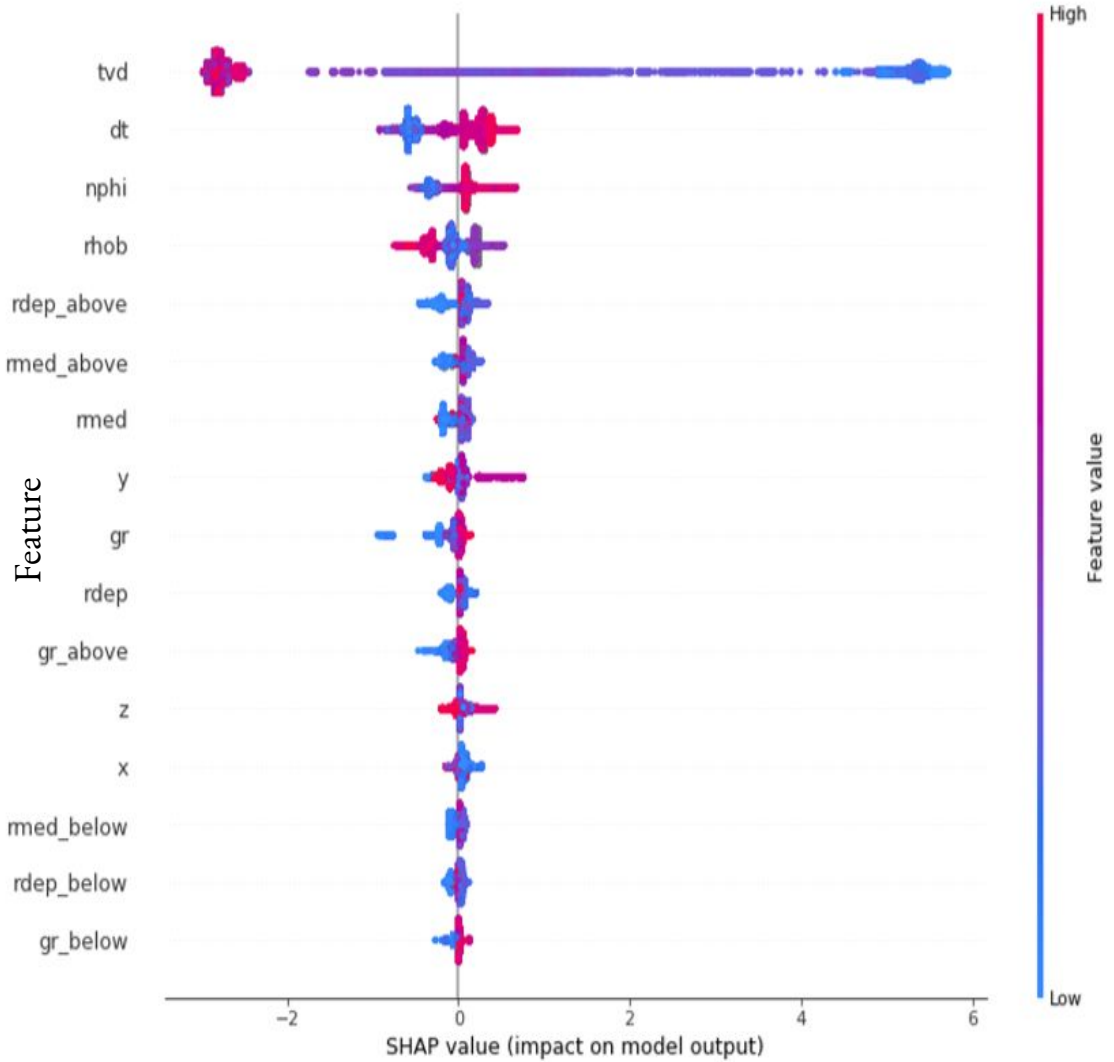
(d) LSTM

SHAP Results (SHapley Additive exPlanations) - LightGBM 1

- Impact of individual feature on the model output (all formations)



- Impact of all features on a specific formation (Formation ‘A’)
- In this case, low TVD value increases the probability of the formation being classified as ‘A’.



Results

- .

	LGBM 1	
	Formation	Group
16/3-7	0.7132	0.945
16/2-14 T2	0.583	0.9153
16/2-11 A	0.8569	0.9472
16/3-6	0.8109	0.9332
16/2-7 A	0.888	0.9673
Total	0.7758	0.9423

Validation Score : 79.06 %

Generalised

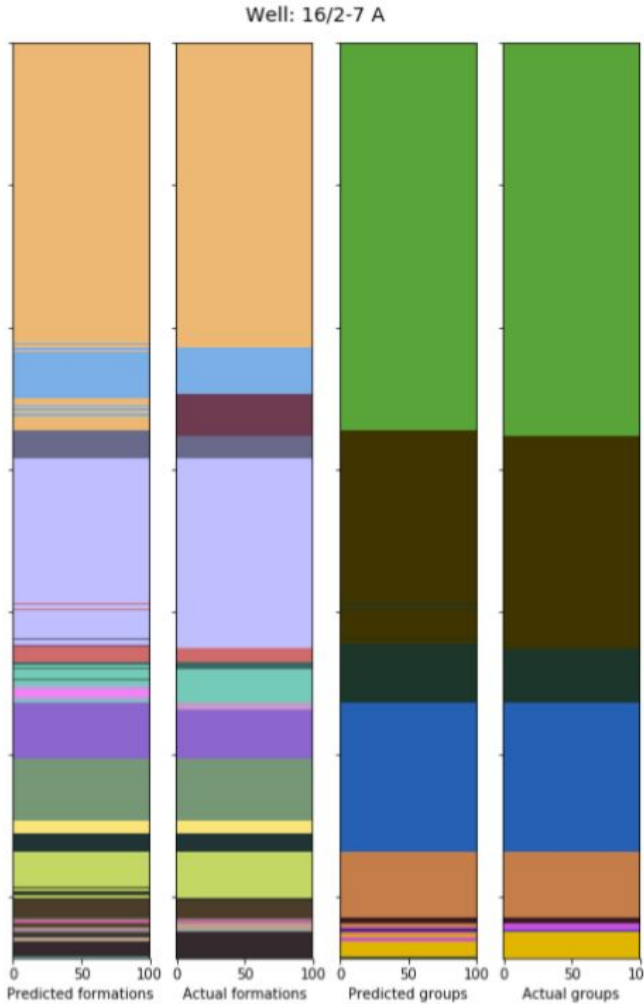
	LGBM Stratified	
	Formation	Group
16/3-7	0.7091	0.9616
16/2-14 T2	0.5698	0.9451
16/2-11 A	0.8698	0.9651
16/3-6	0.8637	0.9559
16/2-7 A	0.9184	0.9807
Total	0.7917	0.9621

Validation Score : 99.06 %

Overfitting?

Conclusion

- Transition points are hard to predict
- Correct group is predicted remarkably
- Architecture of LightGBM 1 model can be reused to predict formations in other area
- Investigate LSTM model in detail
- Model Ensembling
- Make the model scalable



(b) LightGBM 1

THANK YOU

Points to Remember

- Transition points are hard to predict
- Correct group is predicted remarkably
-

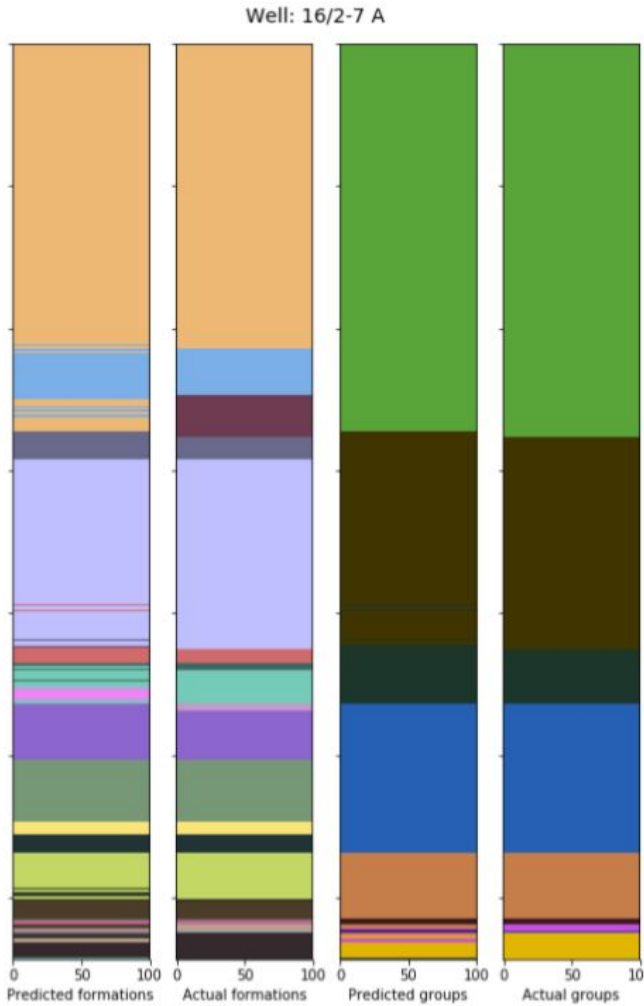
Points to Remember

- The Shapley value, coined by Shapley (1953)41, is a method for assigning payouts to players depending on their contribution to the total payout.
-

- Semi public data
- Quantity of data - how many wellbores - logs with wellbores
- Explain logs, stratigraphy and its use case
- Selected an area
- Logistic Regression - Base Model - Results
- LightGBM 2 - Feature Engineering - Bayesian Optimization - Hyperparameters - Results - SHAP Results
- LightGBM 1 - Feature Engineering - BO - Results - SHAP 77 - 43. Different Stratified groupby shuffle split solution
- LSTM

Conclusion

- Transition points are hard to predict
- Correct group is predicted remarkably
- Architecture of LightGBM 1 model can be reused to predict formations in other area

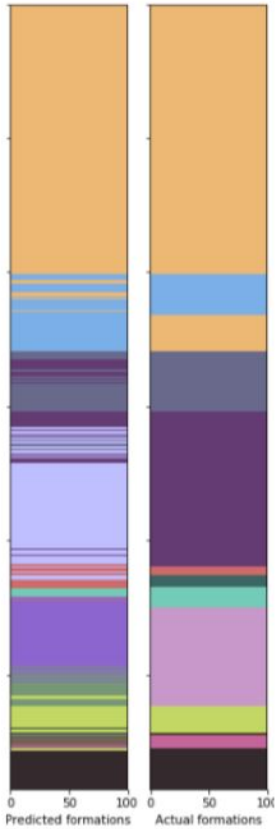


(b) LightGBM 1

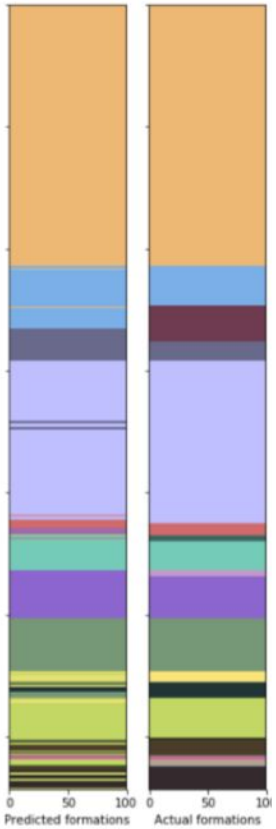
Results - Logistic Regression

- Formations were predicted using the model
- Predicted formation was then correlated with the group it belongs to

	LogReg	
	Formation	**Group**
16/3-7	0.692	0.9312
16/2-14 T2	0.5846	0.956
16/2-11 A	0.6317	0.9384
16/3-6	0.7656	0.9062
16/2-7 A	0.8592	0.9415
Total	0.7063	0.9347



(a) Well 16/2-14 T2



(b) Well 16/2-7 A

Conclusion / Way Forward

- Investigate LSTM model in detail
- Model Ensembling
- Make the model scalable

Workflow

- Machine Learning Workflow