Machine Learning in Oil and Gas

Submitted in Partial Fulfilment for the B. Tech Core Course on

Chemical Engineering Capstone Project (CP302)

By
Mohamed Mazhar Laljee (2019CHB1051)

Department of Chemical Engineering

Under the Guidance of *Dr. Asad H. Sahir*

Indian Institute of Technology, Ropar
Rupnagar, Punjab -14001
May 2022

Acknowledgement

I would like to express a deep sense of thanks and gratitude to my project advisor, Dr. Asad H. Sahir. This would not have been possible without him, as all through the course, he has guided me and corrected me at every step. His constructive criticism and constant drive contributed to the success of this project. I would also like to thank Dr. Chandi Sasmal and Dr. Himanshu Paliwal for the extremely valuable feedback they provided during the mid-semester evaluation. The guidance and resources provided by Dr. Manas Pathak, CEO, EarthEn Inc. and Dr. Palash Panja, CTO, EarthEn Inc. too was instrumental in the success of this project.

Last but not the least, I would like to thank all those who have helped directly or indirectly towards the completion of this project.

Abstract

Traditionally, analytical and numerical methods have been used to model oil and gas exploration and production (E&P) activities. However, these methods have their own disadvantages to which machine learning (ML) provides a solution. With the help of ML, data-driven models can be created using limited data to gain quality insights. In this study, chiefly two ML algorithms have been used- random forest regression and long short-term memory (LSTM) to create an end-to-end workflow which assists crucial E&P activities. The study is based on open-source data from the Volve oilfield released by Equinor. Continuous learning models have been developed for rate of penetration prediction in well drilling and oil and gas production forecasting where both subsurface and over-ground parameters influence the results. On the other hand, a sonic log prediction model is created using reference wells and validated using blind data, where the results are influenced by subsurface parameters only. These models perform satisfactorily, with the R^2 ranging from 0.54 to 0.85. Despite the apparently low R^2 values, the models do a pretty good job in capturing the information and insights that can be gained from the data.

Keywords: exploration and production, machine learning, data-driven models, random forest regression, long short-term memory, rate of penetration, sonic log

Table of Contents

List of Figures

List of Tables

1. Introduction

Accurate prediction of subsurface phenomena is challenging but at the same time essential for the efficient development and management of equipment and oil and gas resources. Oil and gas contribute to about 58% of the global energy mix *(1)*, which makes the activity of collating subsurface and operational data via measurements and computational/mathematical model predictions crucial to help engineers perform economic evaluations and optimization of operational routines. However, this activity is severely complicated due to reservoir heterogeneity, non-linear interdependencies of petrophysical parameters and limited availability of data. Historically, analytical and numerical methods have been used extensively. Analytical methods, although easy to use, make various simplifying assumptions and hence are unable to capture the characteristics specific to a given reservoir. On the other hand, numerical simulations which provide very good results require enormous amounts of data to establish the geological model, numerical model and history matching. This makes them tedious and time-consuming *(1)*. Thus, it is important to strike a middle ground where insights can be derived from limited data without ignoring the innate complexities of the formation under consideration, and circumventing the need for carrying out tedious modelling and simulations using expensive computational resources at the same time. This is why the application of Machine Learning (ML) in the Oil and Gas Industry is gaining widespread interest. ML models can be deployed for a range of exploration and production (E&P) activities such as reservoir characterization, well drilling, forecasting, maintenance, etc.

Upon identification of a suitable reservoir, oil and gas wells are drilled to produce hydrocarbons. When drilling oil and gas wells, time is one of the most important parameters affecting the cost. The drilling rate or rate of penetration (ROP) can be maximized to save costs. The ROP is affected by the selection of the drill bit and drilling mud, and parameters like weight on bit (WOB), revolutions per minute (RPM), formation properties, etc. some of which are controllable while others are not. Selecting an optimal combination of controllable parameters when given a set of uncontrollable parameters can maximize the rate of penetration and reduce drilling time, and thus reduce drilling costs. Therefore, it is crucial to understand the effect of these parameters on the ROP *(2)*. A prerequisite to optimization of the process is the

development of a model which can capture the relationship between the various parameters and ROP to make predictions. Such models can be based on reference wells which are similar to the one being drilled, or can continuously learn from the data measured on-the-go during drilling.

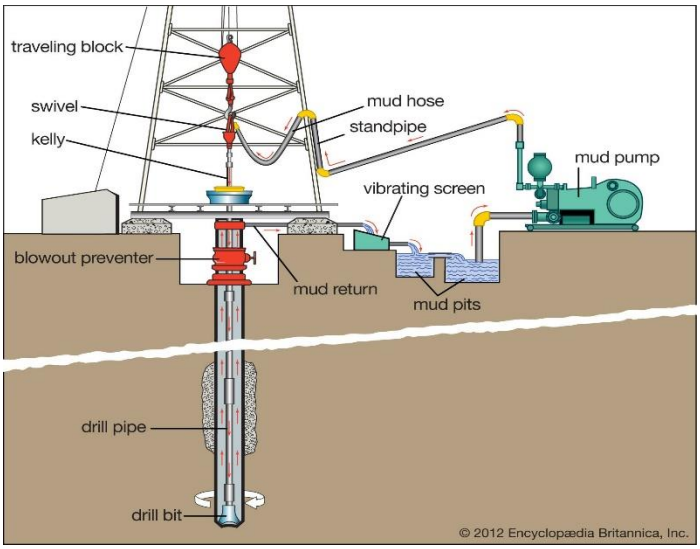

Fig. 1: Oil Well Drilling

Source: ***(3)***

The petrophysical and geomechanical rock properties are essential for studying the characteristics of formations. Tools available for the same include seismic data and well logs. Among well logs, sonic logs (compression and shear wave travel times) are especially important as they give important rock mechanical properties, and an indication of the lithology and fluid saturation, which are commonly used for design calculations in rock fracturing, well development, and production *(4)*. Despite their importance, they seldom exist due to the high costs involved in acquiring them. However, if sonic logs exist for some of the wells in a given reservoir, ML models can be developed to predict the same based on the data obtained during well drilling and logging. The models can then be applied to blind datasets (wells for which sonic logs haven't been acquired).

Forecasting production from a well plays a significant role in a well's life cycle which includes management of resources, adjusting the technology and enhancing recovery. The production data is in the form of a time-series and is

affected by several internal and external factors, making the forecasting exercise complicated and challenging. The decline curve analysis (DCA) method has been widely used; however, it is not robust as important operational parameters such as choke size, wellhead pressure, etc. are not considered. Also, as discussed earlier, numerical methods are laborious *(5)*. Advances in computation and data analytics have presented data driven models as prospective solutions which use measured data as a representation of the 'actual physics' of the system. The forecasts aid cash flow calculations and decision making. Due to the importance of this problem, there is an increasing focus on deploying sophisticated ML techniques such as recurrent neural networks (RNNs) and long-short term memory (LSTM) networks.

The author has not come across any research paper that discusses end-to-end ML modelling of major activities spanning from well drilling to hydrocarbon production. Also, many papers discuss models which use a random training-testing split on the dataset, which in principle, is a flawed methodology as it allows the model to learn from data which must be withheld *(6)*. Such models have limited applicability in the field where continuous learning scenarios are encountered and predictions often have to be made on blind datasets. Thus, this project aims to develop ML models which cover the prediction of ROP for well drilling, prediction of sonic logs for subsurface characterization and hydrocarbon production forecasting for decline curve analysis while trying to emulate as closely as possible the real-world scenarios an engineer might come across in the field.

2. Theory

Before moving to the problems this study intends to solve, it is necessary to briefly discuss the ML techniques that have been used.

2.1. Random Forest Regression

The random forest algorithm is a supervised learning algorithm which uses an ensemble of decision trees to make classification or regression predictions. Here, we are interested in the latter. It has high accuracy, is scalable and easy to use. The algorithm builds *n* decision tree regressors/estimators which are built based on the specification of hyperparameters such as maximum depth, minimum number of samples at the leaf nodes, etc. Each tree trains on a random subset of the data and the sampling is done with replacement. Then, each tree predicts an output for a given input and an average of all those predictions is the model's final output.

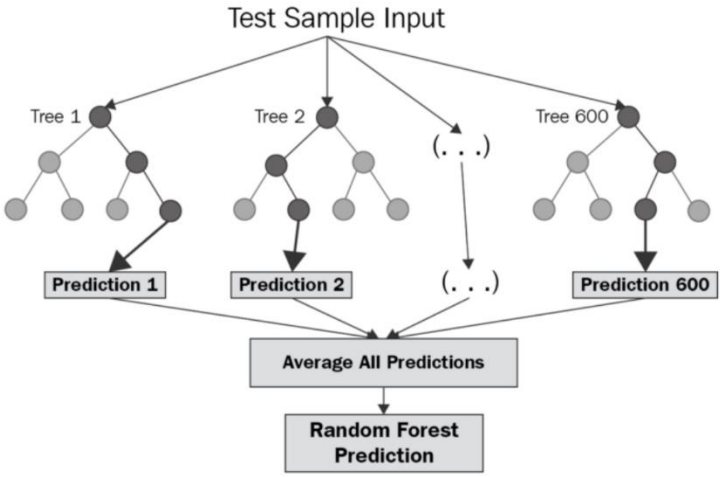

Fig. 2: Random Forest Algorithm
Source: *(7)*

The algorithm is based on the following pattern:

- Attribute selection: The decision tree regression algorithm examines a random subset of attributes and their values in order to identify which attribute value will result in the 'best split.' For regression, the algorithm's objective or cost function is the mean squared error (MSE), which must

be minimized. This is equal to variance reduction as a feature selection criterion.

- When it discovers the best split point candidate, it splits the dataset at that value (known as the root node) and repeats the attribute selection procedure for the other ranges.

- The algorithm continues iteratively until each leaf node has only one sample, or the maximum tree depth or minimum number of samples for a leaf node has been reached *(7)*.

2.2. Long Short-Term Memory (LSTM)

LSTM is an advanced recurrent neural network (RNN) which is sequential in nature, allows past information to persist for a long time and overcomes the problem of vanishing gradient encountered in RNNs. An LSTM cell has 3 parts- a forget gate, an input gate and an output gate.

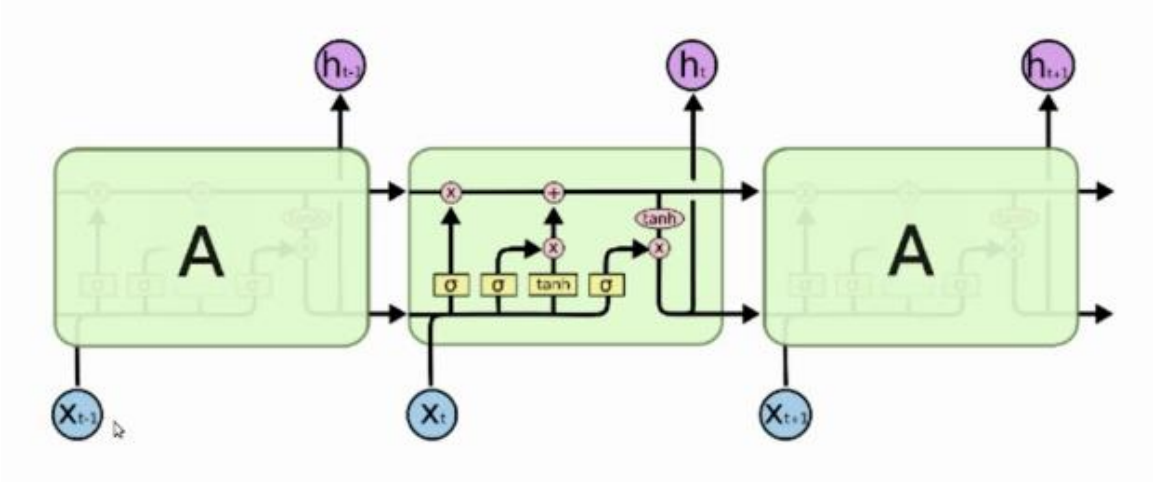

Fig. 3: LSTM Schematic

Source: *(8)*

- Forget gate: This gate decides whether the information coming from the previous timestamp is to be retained or forgotten.

$$f_t = \sigma\left(W_f x_t + U_f H_{t-1} + b_f\right) \qquad (1)$$

Here, x_t is the input to the current timestamp, W_f is the weight associated with the input, H_{t-1} is the hidden state of the previous timestamp, U_f is the weight associated with the hidden state, b_f is the forget gate bias, and σ is the sigmoid function. If $f_t = 0$, everything is forgotten while everything is retained if $f_t = 1$.

- Input gate: This gate quantifies the importance of new information carried by the input.

$$i_t = \sigma(W_i x_t + U_i H_{t-1} + b_i) \qquad (2)$$

Here, W_i is the weight associated with the input, U_i is the weight associated with the hidden state, and b_i is the input gate bias.

- Output gate: This gate determines the output of the cell.

$$o_t = \sigma(W_o x_t + U_o H_{t-1} + b_o) \qquad (3)$$

Here, W_o is the weight associated with the input, U_o is the weight associated with the hidden state, and b_o is the output gate bias.

After the gates have performed their calculation, the current hidden state is calculated as

$$H_t = o_t * tanh(c_t) \qquad (4)$$

where c_t is the current cell state given by

$$c_t = f_t * c_{t-1} + i_t * \tanh\,(W_c x_t + U_c H_{t-1} + b_c) \qquad (5)$$

where c_{t-1} is the previous cell state, W_c is the weight associated with the input, U_c is the weight associated with the hidden state, and b_c is the output gate bias *(8)*.

3. Dataset

3.1. Data Source- Volve Dataset

The data from the Volve field on the Norwegian Continental Shelf was used in this study. Volve field is a 2 km by 3 km oil-bearing reservoir located between 2750 and 3210 meters below sea level, according to Equinor's field development plan report. It is made up of sandstone and has average properties such as 1000 mD permeability (based on well testing), 0.21 porosity, and 0.93 net-to-gross ratio. The average water saturation of the oil-bearing zone is 0.2. The reservoir pressure and temperature are 340 bar and 110 °C, respectively, at a depth of 3060 m. The field was in operation during the years 2008-16 and produced 63,000,000 barrels of oil in total. The API gravity of crude oil from the Volve field is 29.1°, the specific gravity is 0.881, and the viscosity at 20 °C is 22.5 cSt *(5)*. The data was released to the public by Equinor in May 2018 *(9)*, and has about 40,000 files including seismic data, well log data, reservoir simulation model, etc.

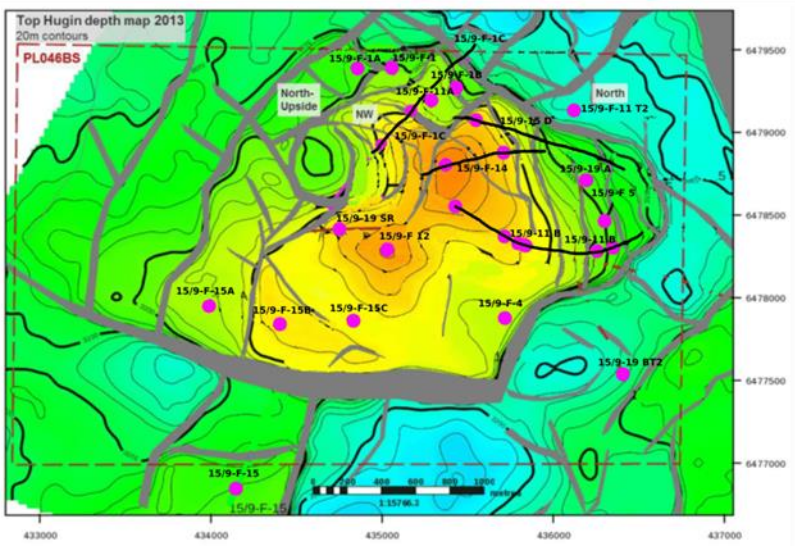

Fig. 4: Contour Map of Volve Oilfield

Source: *(10)*

3.2. Drilling Data

In the Volve dataset repository, the depth-based drilling logs are available in Wellsite Information Transfer Standard Markup Language (WITSML) files which cannot be directly used with Python. Researchers at the University of Stavanger, Norway have parsed these files and made them available in csv format on the university's servers for public access *(11)*. Each parsed file consists of data shown in Table 1.

Table 1: Drilling Data

Attribute	Units
Measured Depth	m
Hole Depth (TVD)	m
Weight on Bit	kkgf
Average Standpipe Pressure	kPa
Average Surface Torque	kN.m
Rate of Penetration	m/h
Average Rotary Speed	rpm
Mud Flow In	L/min
Mud Density In	g/cm^3
Average Hookload	kkgf
Diameter	mm
USROP Gamma	gAPI

For creating the ROP prediction model, data from the deepest well F-15S is used. The data is measured at depths of 1401-4090 m and contains 517,708 rows.

3.3. Sonic Log Data

The sonic log data is available in Digital Log Interchange Standard (DLIS) format in the Volve dataset repository. It has been parsed and made available for public access in Log ASCII Standard (LAS, can be used in Python with the help of lasio library) in a GitHub repository by a geophysicist named Yohanes Nuwara *(12)*. Each parsed file consists of data shown in Table 2.

Table 2: Sonic Log Data

Abbreviation	Attribute	Units
NPHI	Formation porosity	v/v
RHOB	Formation bulk density	g/cm^3
GR	Formation radioactive content	gAPI
RT	Formation true resistivity	ohm-meter
PEF	Formation photoelectric absorption factor	-
CALI	Borehole diameter	inch
DT	Compressional wave travel time	µs/ft
DTS	Shear wave travel time	µs/ft

Well log data for the wells F-11A and F-1B is used for training the model, while that for well F-1A is used as a blind validation set. Satisfactory model performance for the blind set would legitimize field deployment.

3.4. Production Data

The production data of wells is available in csv format in the Volve dataset repository. A single file contains the data for several wells, and consists of attributes shown in Table 3.

Table 3: Production Data

Abbreviation	Attribute	Units
DATEPRD	Date of Record	Date
ON_STREAM_HRS	On Stream Hours	Hours
AVG_DOWNHOLE_PRESSURE	Average Downhole Pressure	bar
AVG_DOWNHOLE_TEMPERATURE	Average Downhole Temperature	^{o}C
AVG_DP_TUBING	Average Differential Pressure of Tubing	bar
AVG_ANNULUS_PRESS	Average Annular Pressure	bar
AVG_CHOKE_SIZE_P	Average Choke Size Percentage	%
AVG_WHP_P	Average Wellhead Pressure	bar
AVG_WHT_P	Average Wellhead Temperature	^{o}C
BORE_OIL_VOL	Oil Volume from Well	m^3
BORE_WAT_VOL	Water Volume from Well	m^3
BORE_GAS_VOL	Gas Volume from Well	m^3

BORE_WI_VOL	Volume of Water Injected	m^3
FLOW_KIND	Type of Flow (Production or Injection)	-
WELL_TYPE	Type of Well (Production or Injection)	-

For the development of a production forecasting model, data from well F-14H (longest producing well) was considered. The set has a total of 3056 rows.

4. Methodology

Although multiple workflows, algorithms and hyperparameters were used to solve the problems, only the best performing methodologies, and the corresponding results will be discussed in the interest of brevity.

4.1. Data Cleaning and Preprocessing

Before creating an ML model, it is necessary to remove outliers, unimportant features and appropriately prepare the data. Exploratory data analysis and domain knowledge are used here.

The ROP dataset contains several measurements for every 1 m of drilled depth. The number of such measurements varies across intervals and has the potential to heavily influence the model's performance based on the conditions encountered in a localized interval of the drilling operation. To circumvent this issue, only one data row per meter of drilled depth is included in the analysis. This also reduces the computational power required. Next, the hole diameter and hole depth/true vertical depth (TVD) features are dropped as the diameter attribute does not affect the drilling rate, while the TVD attribute provides the same information as the measured depth. The data need not be scaled between 0-1 as random forest regression is used and its performance is not affected by the scale and magnitude of the data. Here, outliers cannot be removed at the beginning as a continuous learning model is built.

For the sonic log training and blind validation datasets, the CALI/borehole diameter feature is dropped as the compression and shear wave travel times are influenced by formation properties alone and not something exogenous. Since the formation resistivity and gamma ray measurements are log normally distributed, a logarithmic transformation is applied to obtain a normal distribution which would help in outlier removal and better describing the sonic log data due to an increase in the Pearson correlation coefficient magnitude. Lastly, data rows in which any of the parameters lie outside the mean $\pm$ three times the standard deviation range are considered outliers and removed. Here too the data need not be scaled as random forest regression is used.

The production period for well F-14H lasts from February 2008 to September 2016, however, only the part of the data which represents declining production (July 2013-July 2016) is considered. A 'Days' column is added, where day no. 1 is assigned to the first row in the July 2013-July 2016 subset. Also, not all of the available features are used for the prediction modelling, based on domain knowledge. The features that made the cut include Days, On Stream Hours, Average Downhole Pressure, Average Downhole Temperature, Average Choke Size Percentage, Average Wellhead Pressure, Average Wellhead Temperature, Oil Volume from Well, Gas Volume from Well and Water Volume from Well. Lastly, since an LSTM has been used whose performance is affected by the scale of the data, the data points are scaled to a range of 0-1 using the transformation

$$x_{i,normalized} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \qquad (6)$$

where $x_{i,normalized}$ is the normalized value of a data point x_i under a data type whose minimum and maximum values are given by x_{min} and x_{max} respectively.

4.2. Learning and Prediction Workflows

The continuous learning random forest ROP prediction model learns from the drilling data for the first 30 m and then makes drilling rate predictions for the next 10 m. After the well has been drilled further 10 m deep, the new data is added to the training set and the first 10 data rows are dropped to give a new training set. Predictions are then made for the next 10 m. This can be visualized with the help of Fig. 5.

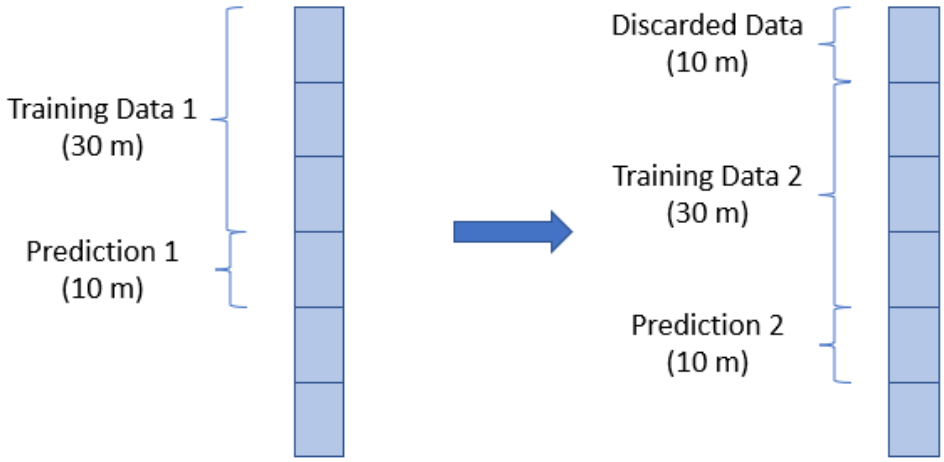

Fig. 5: ROP Prediction Model Workflow

The workflow for the prediction of sonic logs is rather simple, where a random forest regression model is trained on data from wells F-11A and F-1B, and validated on blind data from well F-1A before deployment in the field to predict the logs for new wells.

The workflow for production forecasting using LSTM is rather tricky. Production data from day n to day n + 29 forms a group of points which is used to forecast production for the n + 44th day. Similarly, data for day n + 1 to day n + 30 is used to forecast production for the n + 45th day, and so on. In effect, a group of 30 data rows makes a prediction 15 days into the future. 15 such 'point groups' and targets are used at one time as training data, and the next 15 'point groups' are passed as input to the model to make 15 future predictions. Then, the original training data is replaced by the 15 groups on which predictions were just made, and the corresponding actual target values to form the new training set. This process continues till the dataset is exhausted. This can be visualized with the help of Fig. 6.

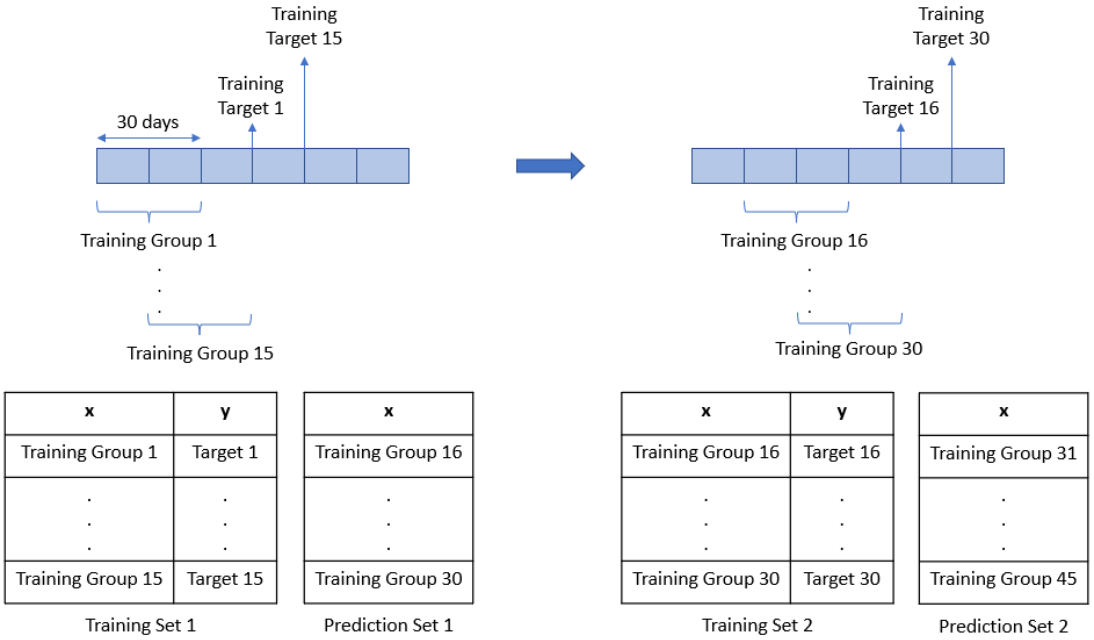

Fig. 6: Production Forecasting Model

4.3. Model specifications

For the ROP prediction model, a random forest regressor model with the specifications shown in Table 4 is created.

Table 4: ROP Prediction Model Specifications

Hyperparameter	Value
n_estimators	500
criterion	'squared error'
max_depth	None
min_samples_split	2
min_samples_leaf	1

For sonic log prediction, a multi-output (to make both compression and shear wave travel time predictions at the same time) random forest regression model is created with the same specifications as in Table 4, except that 100 estimators are used instead of 500. To improve the shear wave travel time prediction, a single-output random forest regression model is created which incorporates the compression wave travel times predicted by the multi-output model as one of the input features. This model has the same specifications as that of the multi-output model.

For well production forecasting, an LSTM having a 10-node input layer with input shape (15, 10) and a hidden layer with 30 nodes- both having sigmoid activation function, is used. The output layer has 2 nodes (one each for oil and gas production). The number of epochs for training is set to 100, and 'adam' optimizer with specifications shown in Table 5 is used.

Table 5: Adam Optimizer Specifications for Production Forecasting

Hyperparameter	Value
Learning rate	0.01
Exponential decay rates for the 1st moment estimates, β_1	0.9
Exponential decay rates for the 2nd moment estimates, β_2	0.999
Numerical stability constant, ε	10^{-7}

5. Results and Discussion

Now that the data-driven models have been established, their performance must be evaluated. Two metrics- coefficient of determination (R^2, indicates goodness of fit) and mean absolute percentage error (MAPE) are used.

The predictions from the continuous learning ROP model are compared with the actual drilling rates. A raw comparison yields an R^2 value of 0.678 and an MAPE value of 20.75%. Upon removing outliers (data points having unusually high or low drilling rates), the R^2 increases to 0.683 and the MAPE reduces to 20.7%. Fig. 7 shows the variation of actual and predicted drilling rates with depth.

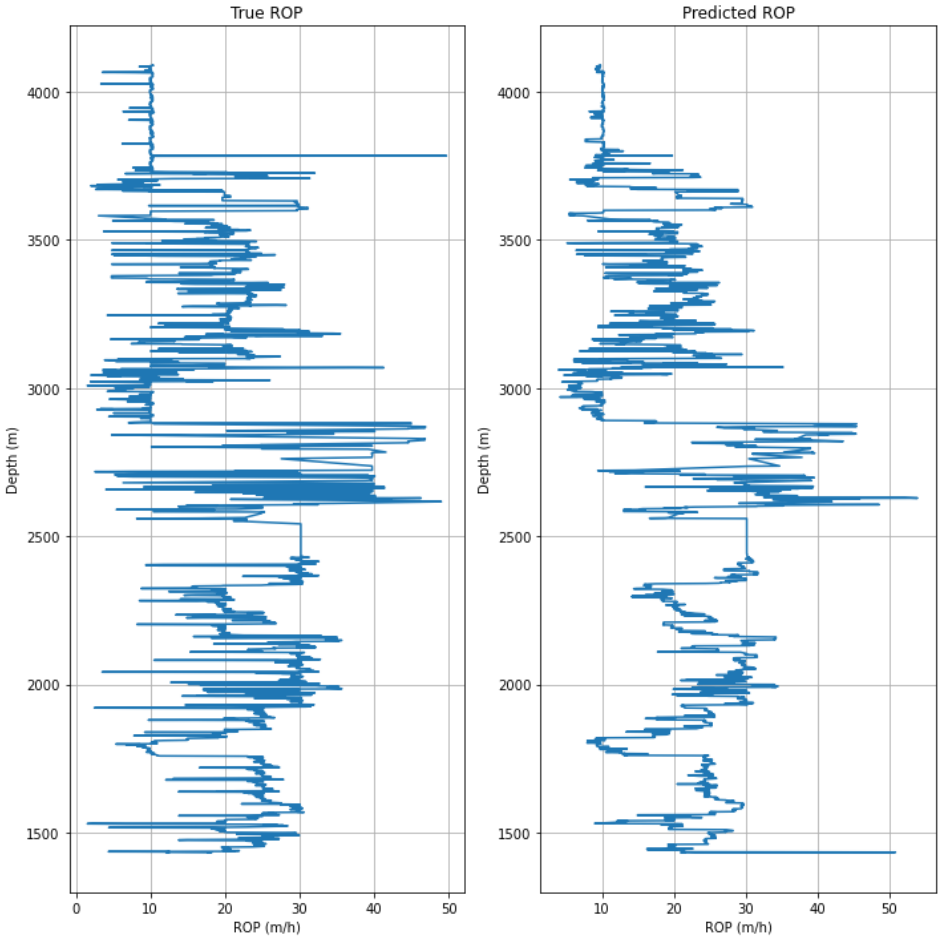

Fig. 7: Actual and Predicted ROP vs Depth

Clearly, the model does a good job in capturing the trend of the data and makes predictions which are mostly close to the real deal, except in cases where the drilling rate suddenly changes by a large amount. The reason for this behavior could be explained by domain experts and drilling staff, and is expected to be difficult to be captured by ML models. Using more sophisticated algorithms such as deep neural networks and incorporating data from other wells could lead to an improvement in the ROP prediction accuracy.

For the sonic log prediction problem, the random forest regressor performs exceptionally well yielding an R^2 value of 0.998 for both the compression and shear wave travel time (DT and DTS respectively). Moving to the blind data, the model predicts the DT values very well with an R^2 of 0.85 and MAPE of 3.92%. Fig. 8 compares the actual and predicted values side by side.

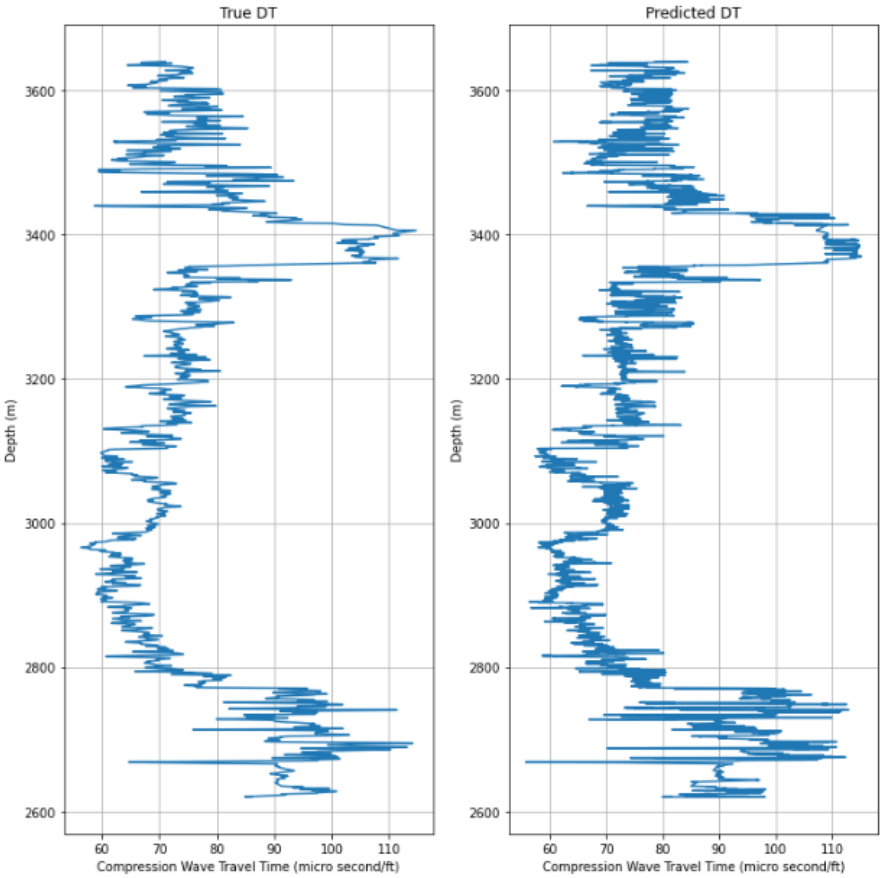

Fig. 8: Actual and Predicted DT vs Depth

However, for DTS, the R^2 is a dismal 0.29. Another random forest regression model is created for DTS prediction where both formation properties and the predicted DT values are used as input features. The R^2 improves by 86% to 0.54 and the MAPE comes out to be an encouraging 6.43%. As can be seen from Fig. 9, the model decently captures the data trend despite a low R^2 and actually performs very well throughout except at lower depths, which probably is the reason behind a low MAPE.

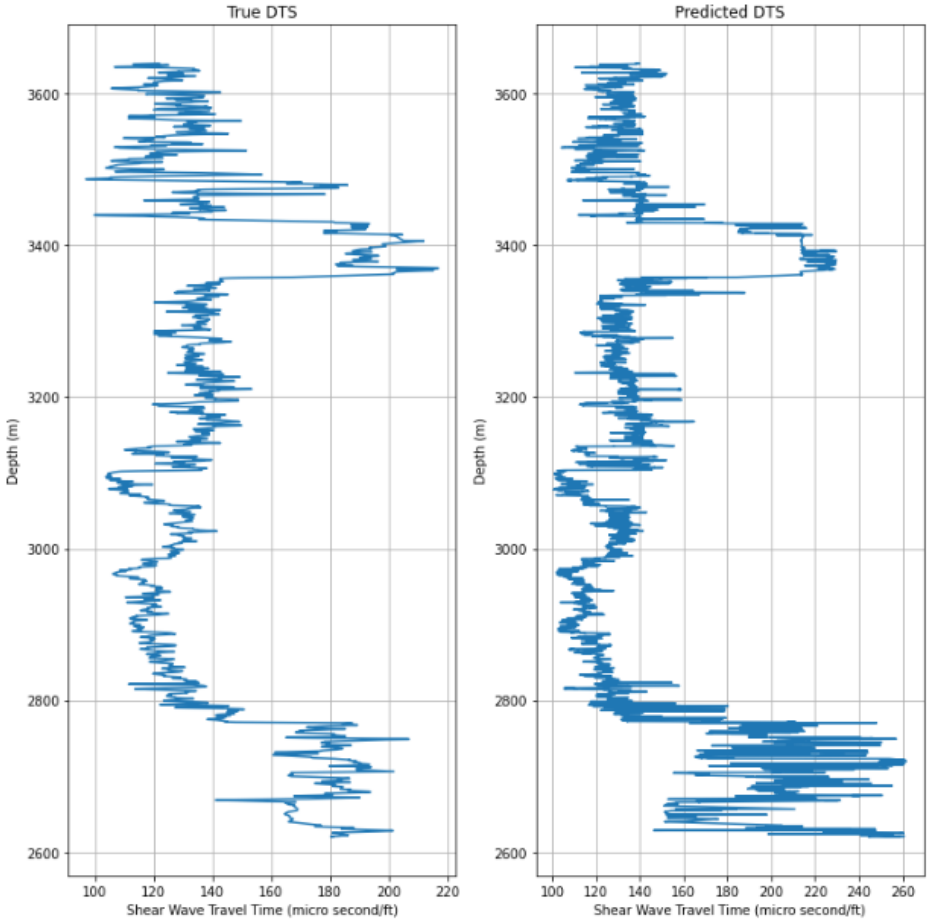

Fig. 9: Actual and Predicted DTS vs Depth

Here too there's an opportunity to increase the prediction accuracy using data from more wells for training and giving more complex ML techniques a shot.

Lastly, the LSTM's performance for predicting the production of both oil and gas is evaluated. After the predictions are made, all data rows for which oil production is less than the volume produced on the last day (68 m^3) are removed and then R^2 and MAPE are calculated. The R^2 for oil production

forecasting is 0.79 and the MAPE is 14.75%. For gas production forecasting, the R^2 and RMSE are 0.76 and 15.5% respectively. Figs. 10 and 11 show that the forecasts and true values are in very good agreement with each other, except where there is a sudden large change in production. Again, domain expertise is needed to decipher such phenomenon, something which is beyond the capacity of data driven models. Here too there is scope for improvement, which could possibly be achieved by increasing the LSTM's complexity.

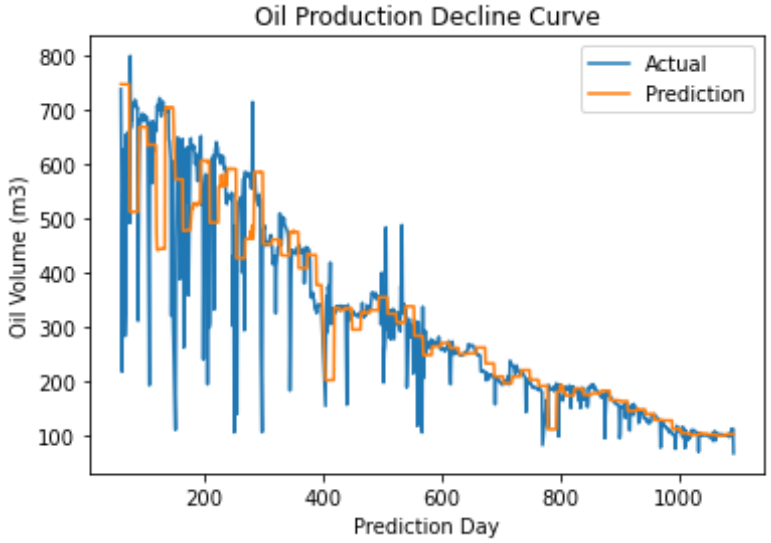

Fig. 10: Comparison of Actual Oil Production and Forecasts

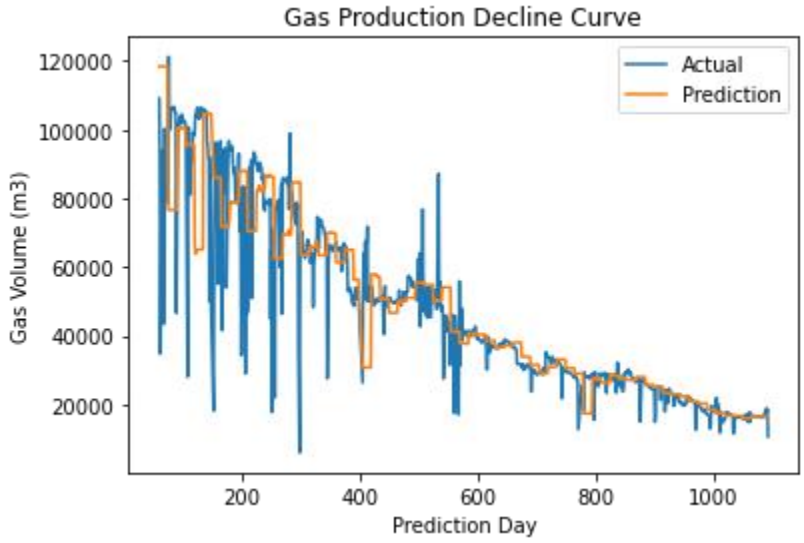

Fig. 11: Comparison of Actual Gas Production and Forecasts

6. Conclusions

In this study, a comprehensive end-to-end machine learning workflow has been developed to support important activities in the upstream oil and gas industry such as well drilling, sonic log prediction and production forecasting. This is important with regards to the fact that oil and gas are limited resources whose production should be carried out in a meticulous, cost effective and environmentally benign manner.

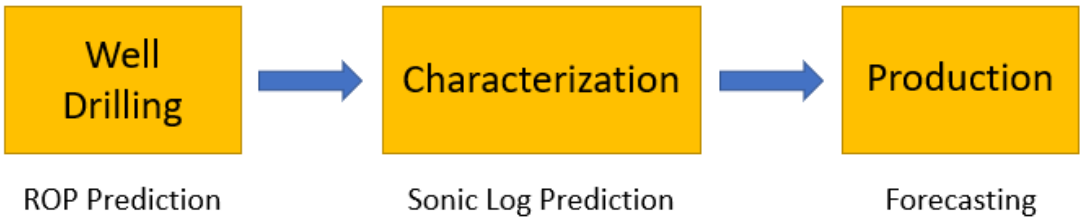

Fig. 12: End-to-End ML Workflow

Vast and complex open-source data from the Volve oilfield was used to understand the challenges faced when working on real-world data. The significance of this work, methodology and ML models have been explained in detail to help make the real-world implementation easy. Figures and tables have been provided wherever possible to aid in understanding of the nature of data and model workflows. The models have been trained and validated to gauge whether they can satisfactorily support production related activities. Further, methods to improve upon the success of this study have also been suggested. In summary, given the lack of a multidisciplinary team of data, reservoir and production engineers, very high-speed computers, funding, structured industrial guidance and experience in the subject, satisfactory results have been obtained and the objective of proposing a viable workflow to help solve industrial problems has nevertheless been achieved.

7. Future Scope

This study has focused on making predictions using standalone algorithms. The possibility of using a combination of algorithms for various subtasks to exploit their strengths can be explored in the future. Further, the target variables can be optimized using algorithms such as Particle Swarm Optimization (PSO) and Genetic Algorithm with the help of prediction models created. It would be interesting to explore in further detail the Volve dataset and solve problems related to well completion, hydraulic fracturing, enhanced oil recovery, etc. Parsing of the data (which usually is in the form used in the oil and gas industry only) could also be undertaken to make it more accessible in csv and excel formats. Data from other fields such as Marcellus shale and Hutton oilfield could also be looked at and incorporated into the training of models to improve accuracy and robustness. Lastly, it would be very interesting and exciting to collaborate with the industry for implementing the models to active producing assets.

8. References

1. Fan, D.; Sun, H.; Yao, J.; Zhang, K.; Yan, X.; Sun, Z. Well production forecasting based on ARIMA-LSTM model considering manual operations. *Energy*. **2021**, *220*, 1. DOI: 10.1016/j.energy.2020.119708 (accessed 2022-4-7)
2. Elkatatny, S. Real-time prediction of rate of penetration while drilling complex lithologies using artificial intelligence techniques. *Ain Shams Engineering Journal*. **2021**, *12(1)*, 917. DOI: 10.1016/j.asej.2020.05.014 (accessed 2022-4-7)
3. Britannica, T. Editors of Encyclopedia. drilling mud. In *Encyclopedia Britannica;* Britannica, 2017 (accessed 2022-4-7)
4. Gamal H.; Alsaihati A.; Elkatatny S. Predicting the Rock Sonic Logs While Drilling by Random Forest and Decision Tree-Based Algorithms. *Journal of Energy Resources Technology*. **2022**, *144(4)*, 1. DOI: 10.1115/1.4051670 (accessed 2022-4-8)
5. Ng, C.S.W.; Ghahfarokhi, A.J.; Amar, M.N. Well production forecast in Volve field: Application of rigorous machine learning techniques and metaheuristic algorithm. *Journal of Petroleum Science and Engineering*. **2022**, *208 Part B*, 1-8. DOI: 10.1016/j.petrol.2021.109468 (accessed 2022-4-8)
6. Tunkiel, A.T.; Sui, D.; Wiktorski, T. Reference dataset for rate of penetration benchmarking. *Journal of Petroleum Science and Engineering*. **2021**, *196*, 8. DOI: 10.1016/j.petrol.2020.108069 (accessed 2022-4-9)
7. The Ultimate Guide to Random Forest Regression. *Keboola*, September 17, 2020. https://www.keboola.com/blog/random-forest-regression (accessed 2022-4-11)
8. Saxena, S. Introduction to Long Short Term Memory (LSTM). *Analytics Vidhya*, March 16, 2021. https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/ (accessed 2022-4-11)
9. Volve field data set. *Equinor*. https://www.equinor.com/energy/volve-data-sharing (accessed 2022-3-10)
10. Ankit. Geological Cross Sections for the Volve Oilfield. *DiscoverVolve*, March 25, 2022. https://discovervolve.com/2022/03/25/geological-cross-sections-for-the-volve-oilfield/ (accessed 2022-4-13)
11. Tunkiel, A. Selected Work Repository. *University of Stavanger*, 2020. https://www.ux.uis.no/~atunkiel/ (accessed 2022-4-10)

12. Nuwara, Y. volve-machine-learning. *GitHub*. https://github.com/yohanesnuwara/volve-machine-learning (accessed 2022-4-20)

9. Checklist

Sr. no.	Task	Status
1	Cover page with proper title, mentor and student names and Institute name and logo	✓
2	Dedication and acknowledgement	✓
3	Executive summary/abstract	✓
4	Table of contents with major sections along with page numbers	✓
5	Each section should start from a new page and have section number	✓
6	Introduction	✓
7	Method	✓
8	Results	✓
9	Discussion	✓
10	Conclusion	✓
11	Future work	✓
12	References	✓
13	Figures and tables should have figure and table numbers and the corresponding captions at the bottom	✓
14	References should be done as per ACS guidelines	✓
15	Plagiarism should be avoided at all costs	✓
16	Use consistent font type, size and justification in the entire document	✓
17	Use Grammarly to check the grammar and sentence structure	✓
18	Use Mendeley or one note to populate the references	✓