An interpretable machine learning methodology for well data integration and sweet spotting identification

Jorge Guevara IBM Research
Bianca Zadrozny IBM Research
Alvaro Buoro IBM Research
Ligang Lu Shell Inc.
John Tolle Shell Inc.
Jan Limbeck Shell Inc.
Mingqi Wu Shell Inc.
Defletf Hohl Shell Inc.

1 Abstract

The huge amount of heterogeneous data provided by the petroleum industry brings opportunities and challenges for applying machine learning methodologies. For instance, petrophysical data recorded in well logs, completions datasets and well-production data also constitute good examples of data for training machine learning models with the aim of automating procedures and giving data-driven solutions to problems arisen in the petroleum industry.

In this work[1], we present a machine learning methodology for oil exploration that 1) opens the possibility of integration of heterogeneous data such as completion, engineering, and well production data, as well as, petrophysical feature estimation from petrophysical data from horizontal and vertical wells; 2) enables the discovery of new locations with high potential for production by using predictive modeling for sweet spotting identification; 3) facilitates the analysis of the effect, role, and impact of some engineering decisions on production by means of interpretable Machine learning modeling, allowing the model validation; 4) allows the incorporation of prior/expert knowledge by using Shape Constraint Additive Models and; 5) enables the construction of hypothetical "what-if" scenarios for production prediction.

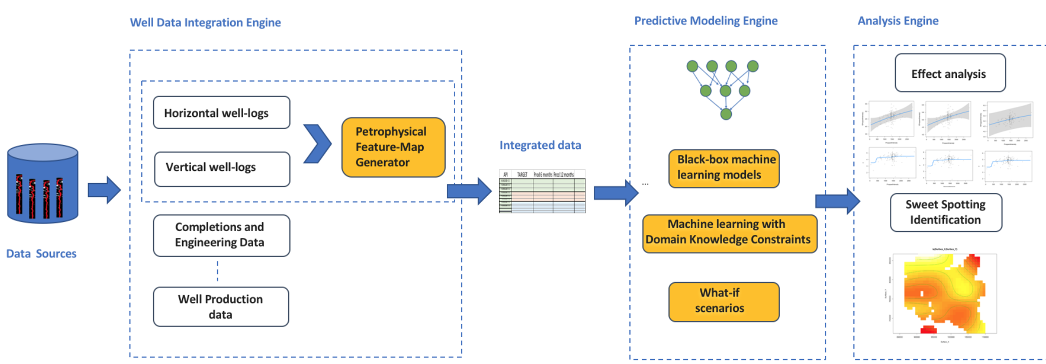

Figure 1: An overview of the proposed methodology.

We depict the overall workflow in Figure 1. Firstly, the *Well Data Integration Engine* processes and integrates all the data from different data sources; (Completions, engineering data, well production data, petrophysical measurements gathered from vertical and horizontal well logs, etc.) We use a

[1]A more detailed version of this work has been published in Guevara et al. (2019)

36th International Conference on Machine Learning (ICML 2019), Long Beach, USA.

Petrophysical Feature Map Generator based on Gaussian Process Regression (Kriging) for characterizing the petrophysical information within a geological layer. Thus, the output of the *Well Data Integration Engine* will be an *integrated dataset*, containing predictors and petrophysical feature values and the cumulative production of oil as the predicted variable.

The next component is the *Predictive Modeling Engine* which uses 1) black-box machine learning models for well production prediction, 2) Generalized Additive Models for model interpretability and 3) Shape Constraint Additive Models (SCAMs) for prior knowledge incorporation. We use all the above models for constructing hypothetical what-if scenarios.

Finally, we perform an effect analysis of predictors of models on the predicted cumulative production by means of conditional plots based on residual plots analysis. We also do sweet spotting identification, i.e., estimate locations with high probability for well exploration and production within a specific geological zone.

We validated this methodology using real well production data (completions, gamma-ray measurements, production data, etc). We used nested leave-one-out cross-validation for assessing the generalization power of models. Among the results, it is important to highlight that 1) performance improves by including prior knowledge via SCAMs, for example, we have a percentage change of 24% between the best Root-mean-squared error (RMSE) result from black-box ML models vs a model that incorporates prior knowledge (SCAMs). 2) we were able to construct hypothetical what-if scenarios based on actual petrophysical data and hypothetical completion and engineering values, 3) we were able to assess the validity of ML models through effect analysis via conditonal plots.

References

Bibor, I. and Szabó, N. P. (2016). Unconventional shale characterization using improved well logging methods. *Geosciences and Engineering*, 5(8):32–50.

Bowman, T. (2010). Direct method for determining organic shale potential from porosity and resistivity logs to identify possible resource plays. In *AAPG Annual Convention, New Orleans, LA*, pages 11–14.

Cawley, G. C. and Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, 11(Jul):2079–2107.

Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York, NY, USA:.

Guevara, J., Kormaksson, M., Zadrozny, B., Lu, L., Tolle, J., Croft, T., Wu, M., Limbeck, J., and Hohl, D. (2017). A data-driven workflow for predicting horizontal well production using vertical well logs. In *DM4OG*, 27-29 April, Houston, Texas, USA.

Guevara, J., Zadrozny, B., Buoro, A., Lu, L., Tolle, J., Limbeck, J., Wu, M., and Hohl, D. (2018). A hybrid data-driven and knowledge-driven methodology for estimating the effect of completion parameters on the cumulative production of horizontal wells. In *SPE Annual Technical Conference and Exhibition*, Paper SPE-191446-MS, 24-26 September, Dallas, Texas, USA. Society of Petroleum Engineers.

Guevara, J., Zadrozny, B., Buoro, A., Lu, L., Tolle, J., Limbeck, J. W., Hohl, D., et al. (2019). A machine-learning methodology using domain-knowledge constraints for well-data integration and well-production prediction. *SPE Reservoir Evaluation & Engineering*.

Hastie, T. J. (2017). Generalized additive models. In *Statistical models in S*, pages 249–307. Routledge.

Kormaksson, M., Vieira, M. R., and Zadrozny, B. (2015). A data driven method for sweet spot identification in shale plays using well log data. In *SPE Digital Energy Conference and Exhibition*, Paper SPE-173455-MS, 3-5 March, The Woodlands, Texas, USA. Society of Petroleum Engineers.

Nunez, G. J., Sandoval Hernandez, G., Stundner, M., Al-Kinani, A., Zangl, G., Mota, M. F., Escalona Gonzalez, H. M., Holy, R., Fuehrer, F., et al. (2010). Integrating data mining and expert knowledge for an artificial lift advisory system. In *SPE intelligent energy conference and exhibition*, Paper SPE-128636-MS, 23-25 March, Utrecht, The Netherlands. Society of Petroleum Engineers.

Pya, N. and Wood, S. N. (2015). Shape constrained additive models. *Statistics and Computing*, 25(3):543–559.

Sharma, R. K. and Chopra, S. (2016). Identification of sweet spots in shale reservoir formations. *first break*, 34(9):43–51.