DATA QUALITY CONSIDERATIONS FOR PETROPHYSICAL MACHINE LEARNING MODELS

Paper Ref : **SPWLA-2021-0036**

Authors : **Andrew McDonald (Lloyd's Register)**

Speaker : **Andrew McDonald**

Abstract:

Decades of subsurface exploration and characterisation have led to the collation and storage of large volumes of well related data. The amount of data gathered daily continues to grow rapidly as technology and recording methods improve. With the increasing adoption of machine learning techniques in the subsurface domain, it is essential that the quality of the input data is carefully considered when working with these tools. If the input data is of poor quality, the impact on precision and accuracy of the prediction can be significant. Consequently, this can impact key decisions about the future of a well or a field.

This study focuses on well log data, which can be highly multi-dimensional, diverse and stored in a variety of file formats. Well log data exhibits key characteristics of Big Data: Volume, Variety, Velocity, Veracity and Value. Well data can include numeric values, text values, waveform data, image arrays, maps, volumes, etc. All of which can be indexed by time or depth in a regular or irregular way. A significant portion of time can be spent gathering data and quality checking it prior to carrying out petrophysical interpretations and applying machine learning models. Well log data can be affected by numerous issues causing a degradation in data quality. These include missing data - ranging from single data points to entire curves; noisy data from tool related issues; borehole washout; processing issues; incorrect environmental corrections; and mislabelled data.

Having vast quantities of data does not mean it can all be passed into a machine learning algorithm with the expectation that the resultant prediction is fit for purpose. It is essential that the most important and relevant data is passed into the model through appropriate feature selection techniques. Not only does this improve the quality of the prediction, it also reduces computational time and can provide a better understanding of how the models reach their conclusion.

This paper reviews data quality issues typically faced by petrophysicists when working with well log data and deploying machine learning models. First, an overview of machine learning and Big Data is covered in relation to petrophysical applications. Secondly, data quality issues commonly faced with well log data are discussed. Thirdly, methods are suggested on how to deal with data issues prior to modelling. Finally, multiple case studies are discussed covering the impacts of data quality on predictive capability

Bio:

Andy McDonald is a Petrophysicist with Lloyd's Register in Aberdeen and has over 15 years of industry experience. He currently provides petrophysical expertise to software development projects and specialises in Python development, artificial intelligence and applications of machine learning to petrophysics. Andy holds an MSc in Earth Science from the Open University, and a BSc (Hons) in Geology & Petroleum Geology from the University of Aberdeen. He has also co-authored several technical conference papers for the SPWLA and SPE on topics covering machine learning, heavy oil, geomechanics and low salinity waterflooding.