

ABSTRACT

Maintaining the production rate of mature oil and gas fields is a big challenge for the oil industry. As one of the common stimulation methods, acidizing repairs formation damage and thus can resume and increase production. Despite the large application of acidizing in oil wells worldwide, the failure of acidizing is still relatively high, especially in sandstone. South Sumatra sandstone acidizing also faces the same challenge with a low success ratio.

We demonstrate a modelling technique to improve the success ratio of matrix acidizing in a hydraulic fractured sandstone formation. Currently, the success rate of matrix acidizing in hydraulic fractured sandstone formation is less than 55%, much lower compared to the more than 90% success rate in carbonate formation. Supervised machine learning with 4 models of the neural network, logistic regression, tree, and random forest was selected to predict the successfulness of matrix acidizing in hydraulic fracturing. In parallel multivariate analysis of principal component regression and partial least square regression approach was utilized to predict the oil gain of the job. Our model has been proposed to overcome the limited availability of solubility and mineralogy data. In conventional preassessment of an acidizing method, new data on solubility and mineralogy are needed and unfortunately often not readily available. Thus, we approximate the solubility and mineralogy data from existing well properties data.

In qualitative prediction, the results showed that the random forest was the best model to predict the successfulness of the job with the area under the curve (AUC) of 0.68 and precision of 0.73 in the training model with 70% of the data. Subsequently, the validation test with the rest of the data (30% of the data) gave 0.51 AUC and 0.61 precision. Meanwhile, in quantitative prediction, the net oil gain was evaluated by using principal component regression (PCR) and partial least square regression (PLS-R). The PCR and PLS-R model gave a coefficient of determination (R^2) of 0.22 and 0.35, respectively. The p -value of PLS-R gave 0.047 (95% confidence interval) which indicates that the model is significant.

The results of this work demonstrate the potential application of supervised machine learning, principal component regression, and partial least square regression to improve candidate selection of oil wells for matrix acidizing especially in hydraulic fractured wells with limited design data.



Mempertahankan tingkat produksi ladang minyak dan gas tua merupakan tantangan besar bagi industri minyak. Sebagai salah satu metode stimulasi umum, pengasaman dapat memperbaiki kerusakan formasi dan dengan demikian dapat melanjutkan dan meningkatkan produksi. Meskipun banyak aplikasi pengasaman di sumur minyak di seluruh dunia, kegagalan pengasaman masih relatif tinggi, terutama di lapisan *sandstone*. Pengasaman *sandstone* di Sumatera Selatan juga menghadapi tantangan yang sama dengan rasio keberhasilan yang rendah.

Kami mendemonstrasikan teknik pemodelan untuk meningkatkan rasio keberhasilan pengasaman dalam formasi *sandstone* yang telah mengalami perlakuan rekah hidrolik. Saat ini, tingkat keberhasilan pengasaman dalam formasi *sandstone* rekah hidrolik kurang dari 55%, jauh lebih rendah dibandingkan dengan tingkat keberhasilan lebih dari 90% dalam lapisan karbonat. *Supervised machine learning* dengan 4 model yaitu *neural network*, *logistic regression*, *tree*, dan *random forest* dipilih untuk memprediksi keberhasilan pengasaman matrik yang telah mengalami rekahan hidrolik. Dalam analisis multivariat paralel dari *principal component regression* dan pendekatan *partial least square regression* digunakan untuk memprediksi peningkatan minyak dari pekerjaan yang dilakukan. Model kami telah diusulkan untuk mengatasi keterbatasan ketersediaan data kelarutan dan mineralogi. Dalam penilaian awal metode pengasaman konvensional, data baru tentang kelarutan dan mineralogi diperlukan dan sayangnya seringkali tidak tersedia. Dengan demikian, kami memperkirakan data kelarutan dan mineralogi dari data properti sumur yang ada.

Dalam prediksi kualitatif, hasil menunjukkan bahwa *random forest* merupakan model terbaik untuk memprediksi keberhasilan pekerjaan dengan area di bawah kurva (AUC) 0,68 dan presisi 0,73 pada model pelatihan dengan 70% data. Selanjutnya, uji validasi dengan sisa data (30% data) memberikan AUC 0,51 dan presisi 0,61. Sementara itu, dalam prediksi kuantitatif, peningkatan minyak dievaluasi dengan menggunakan *principal component regression* (PCR) dan *partial least square regression* (PLS-R). Model PCR dan PLS-R memberikan koefisien determinasi (R^2) masing-masing sebesar 0,22 dan 0,35. Nilai p dari PLS-R memberikan 0,047 (interval kepercayaan 95%) yang menunjukkan bahwa model tersebut signifikan.

Hasil study ini menunjukkan potensi penerapan *supervised machine learning*, *principal component regression*, dan *partial least square regression* untuk meningkatkan pemilihan kandidat sumur minyak yang akan dilakukan pengasaman matriks terutama pada sumur rekah hidrolik dengan data desain yang terbatas.