

ABSTRAK

Penentuan zona produktif reservoir dalam evaluasi petrofisika menggunakan data *well logging* merupakan aspek krusial dalam eksplorasi dan evaluasi hidrokarbon. Metode deterministik memiliki keterbatasan dalam memberikan hasil prediksi karena bergantung pada penetapan nilai *cut-off* empiris. Hal ini menyebabkan risiko kehilangan potensi produksi (*missed opportunities*) ataupun risiko kerugian operasional. Penelitian ini bertujuan untuk membandingkan metode deterministik dan *machine learning* algoritma *Logistic Regression* dalam memprediksi zona produktif pada studi kasus sumur Mangunjaya. Metode deterministik diterapkan melalui analisis *crossplot* multi-parameter untuk menentukan nilai *cut-off* pada Porositas Efektif, Saturasi Air, dan Volume Serpih yang diintegrasikan dengan log *quad-combo*. Pengembangan model *Logistic Regression* mencakup tahapan persiapan data, penggunaan *RobustScaler* pada proses standarisasi, implementasi *Polynomial Features*, serta metode ADASYN dalam menangani ketidakseimbangan data *logging*. Validasi model dilakukan secara komprehensif menggunakan *confusion matrix* dan analisis kurva ROC untuk mengukur performa prediksi zona *pay*. Hasil penelitian ini menunjukkan bahwa metode deterministik cenderung menghasilkan prediksi *underestimation* zona produktif akibat tingginya *false negative*, dengan prediksi 25.36% interval kedalaman sebagai zona *pay*. Sebaliknya, model dengan algoritma *Logistic Regression* menunjukkan performa klasifikasi yang sangat baik dengan akurasi 94.5% dan nilai ROC-AUC sebesar 0.973. Validasi terhadap data aktual (*ground truth*) menunjukkan bahwa metode deterministik memberikan hasil *underestimation* dengan hanya memprediksi sebanyak 555 sampel dari total 1053 zona produktif pada data aktual, sementara model *machine learning* memberikan hasil prediksi sebanyak 1176 sampel produktif. Penelitian ini menggunakan penetapan ambang batas (*threshold*) optimal sebesar 0,7970 yang terbukti mampu meningkatkan objektivitas prediksi dan menangani bias interpretasi pada metode deterministik. Temuan ini menegaskan bahwa algoritma *Logistic Regression* mampu meningkatkan akurasi dan objektivitas prediksi zona produktif dibandingkan metode deterministik berbasis *cut-off*, sehingga menawarkan solusi yang andal untuk mendukung keputusan eksplorasi hidrokarbon pada industri minyak dan gas.

Kata Kunci: Reservoir; Zona Produktif; Metode Deterministik; *Machine Learning*; *Logistic Regression*

ABSTRACT

Determining productive zones in reservoirs in petrophysical evaluations using well logging data is a crucial aspect of hydrocarbon exploration and evaluation. Deterministic methods have limitations in providing predictive results because they depend on empirical cut-off values. This leads to the risk of missing production potential (missed opportunities) or operational losses. This study aims to compare deterministic methods and the Logistic Regression machine learning algorithm in predicting productive zones in the Mangunjaya well case study. The deterministic method is applied through multi-parameter crossplot analysis to determine the cut-off values for Effective Porosity, Water Saturation, and Shale Volume integrated with quad-combo logs. The development of the Logistic Regression model includes data preparation, the use of RobustScaler in the standardization process, the implementation of Polynomial Features, and the ADASYN method in handling data logging imbalances. Model validation was carried out comprehensively using a confusion matrix and ROC curve analysis to measure the performance of pay zone predictions. The results of this study show that the deterministic method tends to produce underestimation predictions of productive zones due to high false negatives, with a prediction of 25.36% depth interval as a pay zone. In contrast, the model with the Logistic Regression algorithm shows excellent classification performance with 94.5% accuracy and an ROC-AUC value of 0.973. Validation against actual data (ground truth) shows that the deterministic method provides underestimation results by predicting only 555 samples from a total of 1053 productive zones in the actual data, while the machine learning model provides prediction results of 1176 productive samples. This study uses an optimal threshold of 0.7970, which has been proven to improve the objectivity of predictions and address interpretation bias in deterministic methods. These findings confirm that the Logistic Regression algorithm can improve the accuracy and objectivity of productive zone predictions compared to deterministic methods based on cut-offs, thus offering a reliable solution to support hydrocarbon exploration decisions in the oil and gas industry.

Keywords: *Reservoir; Productive Zones; Deterministic; Machine Learning; Logistic Regression*