

**MACHINE LEARNING-BASED PREDICTION OF INTERFACIAL TENSION IN WAXY OIL-  
SURFACTANT-WATER SYSTEMS: A STUDY USING LIMITED DATA**

**BACHELOR THESIS**

Hafa Fannan Nugraha

12219033

Submitted as partial fulfillment of the requirements for the degree of

**BACHELOR OF ENGINEERING**

in the Petroleum Engineering study program



PETROLEUM ENGINEERING STUDY PROGRAM  
FACULTY OF MINING AND PETROLEUM ENGINEERING  
INSTITUT TEKNOLOGI BANDUNG

2023

**MACHINE LEARNING-BASED PREDICTION OF INTERFACIAL TENSION IN WAXY OIL-  
SURFACTANT-WATER SYSTEMS: A STUDY USING LIMITED DATA**

**BACHELOR THESIS**

Hafa Fannan Nugraha  
12219033

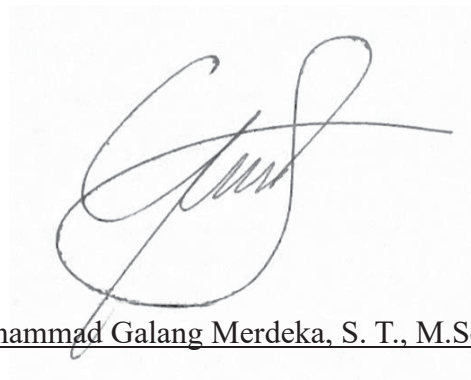
Submitted as partial fulfillment of the requirements for the degree of  
**BACHELOR OF ENGINEERING**  
in the Petroleum Engineering study program

Approved by:  
Thesis Advisor,



Dr. Rani Kurnia, S.Si., M.T.  
NIP. 119110057

Approved by:  
Thesis Advisor,



Mohammad Galang Merdeka, S. T., M.Sc.

# MACHINE LEARNING-BASED PREDICTION OF INTERFACIAL TENSION IN WAXY OIL-SURFACTANT-WATER SYSTEMS: A STUDY USING LIMITED DATA

Hafa Fannan Nugraha<sup>1</sup>, Rani Kurnia<sup>2</sup>, Mohammad Galang Merdeka<sup>2</sup>

Copyright 2023, Institut Teknologi Bandung

## Abstract

Accurately predicting interfacial tension (IFT) in waxy oil-surfactant-water systems is important for optimizing enhanced oil recovery (EOR) processes. However, the limited data availability poses a significant challenge in developing accurate prediction models. This study proposes a machine learning approach that uses the Random Forest model to predict IFT in waxy oil-surfactant-water systems using a dataset with limited samples. The Random Forest algorithm excels at capturing non-linear relationships between input and output variables. The investigation extracts relevant information from the dataset by implementing feature engineering methodologies, develops an effective forecasting model, and understands the principal factors affecting the waxy oil-surfactant interfacial tension prediction. By training the Random Forest model on the limited dataset, this study aims to achieve precise predictions of IFT values. The adaptability and user-friendly nature of the Random Forest algorithm have contributed to its widespread adoption, addressing challenges related to categorization and prediction. The study successfully developed a reliable model, with a test data prediction  $R^2$  value of 0.824, demonstrating its capability to generalize from limited data and accurately predict IFT. These findings significantly enhance oil recovery by providing a reliable and efficient tool for estimating interfacial tension in waxy oil-surfactant-water systems. Simulation results highlight the influence of various parameters on interfacial tension prediction, including Concentration of Surfactant, Oil Solubilization Ratio, Salinity of Surfactant, HLB, and Hydrophobic Number. These results highlight the potential of machine learning techniques, specifically the Random Forest model, in overcoming data limitations and improving prediction accuracy. The results of this study may be beneficial in optimizing surfactant selection, injection strategies, and decision-making processes for enhanced oil recovery applications.

**Keywords:** *interfacial tension, surfactant, machine learning, random forest, prediction*

## Sari

Memprediksi dengan akurat tegangan antarmuka (IFT) dalam sistem minyak lilin-surfaktan-air memiliki kepentingan yang besar untuk mengoptimalkan proses peningkatan perolehan minyak. Namun, keterbatasan ketersediaan data menjadi tantangan yang signifikan dalam mengembangkan model prediksi yang akurat. Penelitian ini mengusulkan pendekatan pembelajaran mesin yang menggunakan model Random Forest untuk memprediksi IFT dalam sistem minyak waxy-surfaktan-air. Algoritma Random Forest sangat baik dalam menangkap hubungan non-linear antara variabel input dan output. Dengan menerapkan metodologi rekayasa fitur, penelitian ini mengekstraksi informasi relevan dari dataset, mengembangkan model prediksi yang efektif, dan memahami faktor-faktor utama yang mempengaruhi prediksi tegangan antarmuka minyak lilin-surfaktan. Dengan melatih model Random Forest pada dataset terbatas, penelitian ini bertujuan untuk mencapai prediksi nilai IFT yang akurat. Adaptabilitas dan sifat yang mudah digunakan dari algoritma Random Forest telah berkontribusi pada adopsi yang luas, mengatasi tantangan yang terkait dengan kategorisasi dan prediksi. Penelitian ini berhasil mengembangkan model yang dapat diandalkan, dengan skor prediksi  $R^2$  pada data uji sebesar 0,824, yang menunjukkan kemampuannya untuk generalisasi dari data terbatas dan memprediksi IFT secara akurat. Temuan ini memberikan kontribusi signifikan dalam pemulihan minyak ditingkatkan dengan menyediakan alat yang dapat diandalkan dan efisien untuk memperkirakan tegangan antarmuka dalam sistem minyak lilin-surfaktan-air. Hasil simulasi menunjukkan pengaruh berbagai parameter terhadap prediksi tegangan antarmuka, termasuk Konsentrasi Surfaktan, dan Rasio Solubilisasi Minyak, Salinitas Surfaktan, HLB, dan Angka Hidrofobik. Temuan ini menekankan potensi teknik pembelajaran mesin, khususnya model Random Forest, dalam mengatasi keterbatasan data dan meningkatkan akurasi prediksi. Wawasan yang diperoleh dari penelitian ini



*dapat membantu dalam mengoptimalkan pemilihan surfaktan, strategi injeksi, dan proses pengambilan keputusan secara keseluruhan untuk aplikasi pemulihan minyak ditingkatkan.*

*Kata Kunci: tegangan antarmuka, surfaktan, machine learning, random forest, prediksi*

---

<sup>1</sup>Student of the Petroleum Engineering Study Program, Institut Teknologi Bandung, Class of 2023

<sup>2</sup>Thesis Adviser in Petroleum Engineering Study Program, Institut Teknologi Bandung