

ABSTRACT

Automated Classification and Information Extraction in Daily Drilling Reports Using Rule-Based Approaches and Named Entity Recognition

by

Hidayah Nurul Hasanah Zen

23/512827/PPA/06516

The planning of proposed wells in the oil and gas industry requires mapping drilling hazards that have occurred in other wells within a radius of approximately 2 KM from the proposed well coordinates. The documents used for drilling hazard mapping are *Daily Drilling Reports* (DDRs). DDRs are detailed reports on daily drilling operations (spanning several months) produced by drilling operators during the drilling of oil and gas wells. Typically, engineers manually extract information from DDRs, which are presented in unstructured free-text formats. This manual process often limits the number of wells evaluated or leads to the evaluation not being conducted at all. Additionally, project planners may spend 60-80% of the total project time solely on data collection.

This study proposes a rule-based approach combined with Named Entity Recognition (NER) to automate the classification of drilling hazard types and the extraction of well names, events, drilling hazard depths, Pore Pressure (PP), and Equivalent Circulating Density (ECD) from DDRs. This automation aims to generate accurate drilling hazard maps faster than manual methods. The proposed method addresses limitations found in previous studies by integrating linguistic rules tailored to DDR technical terminology with NER's capability to understand contextual information.

The results demonstrate that the system successfully performs automatic classification and extraction from DDRs. The average execution time for processing a well and generating a drilling hazard map is less than nine minutes. These findings establish the proposed approach as a state-of-the-art method for event extraction from DDRs. Despite achieving high accuracy, limitations were found in PP/ECD extraction due to variations in DDR text formats not detected by the initial rule set. Therefore, the system requires updates to recognize a broader range of DDR formats.

Keywords: Daily Drilling Report, Named Entity Recognition, automated extraction, drilling hazard classification.

INTISARI

Otomatisasi Klasifikasi dan Ekstraksi Informasi pada *Daily Drilling Reports* Menggunakan Pendekatan Berbasis Aturan dan *Named Entity Recognition*

Oleh

Hidayah Nurul Hasanah Zen

23/512827/PPA/06516

Perencanaan sumur usulan dalam industri minyak dan gas bumi (Migas) memerlukan pemetaan permasalahan pemboran berupa *drilling hazard map* yang pernah terjadi pada sumur lain dengan radius kurang lebih 2 KM dari koordinat sumur usulan. Dokumen yang digunakan dalam pemetaan *drilling hazard* adalah *Daily Drilling Reports* (DDRs). DDRs berupa dokumen rinci mengenai operasi sehari-hari (beberapa bulan) yang dihasilkan oleh operator pemboran selama pemboran sumur Migas. Umumnya, *engineers* mengekstrak informasi pada DDRs dalam format teks bebas dan tidak terstruktur secara manual. Hal ini membuat proses evaluasi sering kali hanya dilakukan pada sejumlah sumur yang lebih sedikit atau bahkan tidak dilakukan sama sekali. Di samping itu, perencana proyek dapat menghabiskan 60-80% dari total waktu proyek hanya untuk mengumpulkan data.

Penelitian ini mengusulkan penggunaan pendekatan berbasis aturan dan *Named Entity Recognition* (NER) untuk mengotomatisasi klasifikasi jenis *drilling hazard* serta ekstraksi *well name*, *event*, kedalaman *drilling hazard*, *Pore Pressure* (PP), dan *Equivalent Circulating Density* (ECD) dari DDRs agar pembuatan *drilling hazard map* tidak hanya lebih cepat dibandingkan metode manual tetapi juga memiliki performa lebih tinggi daripada penelitian sebelumnya, yakni lebih dari 97,3%. Metode ini diharapkan dapat memperbaiki keterbatasan yang ditemukan dalam penelitian sebelumnya dengan mengintegrasikan aturan linguistik yang disesuaikan dengan terminologi teknis DDRs dan kemampuan NER dalam memahami konteks.

Hasil penelitian menunjukkan bahwa sistem berhasil mengklasifikasikan dan mengekstraksi DDRs secara otomatis. Rata-rata waktu eksekusi suatu sumur dalam menghasilkan *drilling hazard map* adalah kurang dari 9 menit. Temuan penelitian ini menetapkan pendekatan yang diusulkan sebagai metode *state-of-the-art* untuk ekstraksi *event* dari DDRs. Meskipun mencapai performa yang tinggi, terdapat keterbatasan pada ekstraksi PP/ECD akibat variasi format teks DDRs yang tidak terdeteksi oleh aturan awal. Oleh karena itu, sistem perlu diperbarui agar dapat mengenali lebih banyak variasi format DDRs.

Kata kunci: *Daily Drilling Report*, *Named Entity Recognition*, ekstraksi otomatis, klasifikasi *drilling hazard*.