Please fill in the name of the event you are preparing this manuscript for.	International Petroleum Technology Conference 2023 (15th IPTC)
Please fill in your 5-digit IPTC manuscript number.	IPTC-23067-Abstract
Please fill in your manuscript title.	Benchmarking of Hook-up and Commissioning activities using Machine Learning pipeline on unstructured data

Please fill in your author name(s) and company affiliation.

Given Name	Surname	Company
Francois	Baillard	Iraya Energies
Nina Marie	Hernandez	Iraya Energies
M Muaz	B Azman	Petronas
Suhaib Moh'd	Taha Yousif	Petronas

Abstract

Hook- up and commissioning (HUC) activities are the last validation before the start-up of an asset for oil and gas production. During this phase, construction errors and design flaws are found and corrected causing expensive project overruns. As an oil and gas asset operator, benchmarking engineering providers across different projects provides an efficient solution and allows to identify trend and variance during project scoping and execution.

The challenge of such a benchmarking exercise is that the necessary data is trapped in internal reports such as close-out and engineering reports with different formats and layouts making the extraction of the information highly manual and time-consuming. In this paper, we are going to demonstrate on how the latest Machine Learning (ML) and Artificial Intelligence (AI) advances automate the searching and extraction of key information for HUC projects benchmarking.

Given the high variability of reports to be analysed, the unstructured data is ingested through a robust and automated ML/AI pipelines that leverages on Natural Language Processing (NLP) and Named Entity Recognition (NER) to identify related information within the text and thus facilitate project's metadata retrieval and extraction such as project name, operator, platform, field, project start date and project end date. In addition, integrating Deep Convolutional Neural Network (DCNN) in the workflow aids the classification of extracted images according to their image classes such as table, drawing and Gantt chart. The processed text and images are then searchable allowing end users to perform any deep search combining extracted text and tagged labels. The last part of the ML/AI pipeline consists of the auto-extraction of activities and the associated parameters for benchmarking purpose easily visualized and exported from the UI/UX front-end.

In this case study, the information from the unstructured data is extracted for nineteen different HUC projects that includes 35,941 processed pages and 25,741 extracted images. During the discovery stage, the deployment of the output of the generic ML/AI enabled the end-user to find the documents having benchmarking information. The speed of search retrieval makes it possible to pin down patterns and format with the information of interest; contained in this case in plain text, tables, and Gantt Chart. Then the expert ML/AI pipeline extracts the different entities: the name of the service providers activities and the associated time to complete the given task. A mapping step standardizes the extraction and provides a systematic extractable output. At any time, a cross-validation can be performed on the automatic extraction to track and confirm the source of the information.

This paper shows the effectiveness of ML and AI technologies to dive into vast amount of HUC unstructured data and extract benchmarking information.