Session Title: S2 : Data Analytics, Data Management & Data Readiness
Session Discipline: Geoscience
Presenter name: Nina Marie Hernandez
Company: Iraya Energies Sdn Bhd
Abstract title: Value Proposition of ElasticDocs Data Management – How can Data Analytics and Machine Learning Techniques Help Companies Reveal Hidden Trends in a New Ventures Effort?

Introduction

When an oil company's New Ventures team moves into a newly acquired portfolio or is investigating a new exploration area, they are usually faced with two problems: 1) having very limited data, or 2.) having lots of it. The first problem is typically dealt with by utilizing geological analogues and applying transfer learning to understand a new region. The second problem is dealt with by optimizing database management systems to make information available to explorationists in the most efficient way possible to generate the much needed fresh perspectives and new geological insights.

An ideal scenario in both cases, would be to have all data in a structured database, but this is rarely the case. To some extent, seismic data and well data will have platforms that index and visualize these datasets. However, many of the information in an exploration area are in the form of unstructured data within documents or archived files.

Method and/or Theory

Here, we propose a novel workflow on unstructured data processing for volumetric geoscience data, which aims to bridge the domain of data management and geoscience data analytics. The ElasticsDocs workflow, is composed of carefully curated set of data analytics and machine learning algorithms, and is a smart pipeline of optical character recognition (OCR), natural language processing (NLP), elasticsearch, unsupervised clustering, and supervised image classification techniques with deep learning.

The appointed algorithms serve different functionalities at every stage of the data exploration and mimics the human experience of mining for geoscience information. OCR allows for extraction of text within documents or images, whereas NLP enables extraction of metadata and understanding of geological context by converting words into vectors.

Examples

The full text search capability within ElasticDocs makes it extremely efficient to browse through thousands of pages of documents in seconds, for important information in the context of New Ventures workflow, for example: which wells have hydrocarbons shows, which have dry holes? What are the hydrocarbon plays within the area and what wells to use as analogues? What are the drilling risks observed, are there many occurrences of overpressure?

Within this workflow, advanced supervised deep learning algorithms have been utilized to recognize frequently used information such as: seismic, cores, thin sections, image logs, rose diagrams among others. Through a novel search-and-cluster technique, the proposed workflow makes it efficient to perform "well twinning" or search for effective geological analogues.

Conclusions

In this paper, we demonstrate how novel machine learning techniques can extract full value out of subsurface databases and reveal previously hidden trends from voluminous text and image data in a New Ventures area, within the shortest time possible.

References

Simonyan, K. & Zisserman, A., 2014: Very Deep Convolutional Networks for Large-scale Image Recognition. *arXiv preprint arXiv:1409.1556*
van der Maaten, L.J.P. & Hinton G.E., 2008]: Visualizing High-Dimensional Data using t-SNE. *Journal of Machine Learning Research*, **9**, 2576-2605.

(Paper Under Review)