L'association Machine Learning
Data Science et Intelligence Artificielle

DATA BATTLE 2023

DATA SCIENCE AS UN ENABLER TO REACH NETZERO

Project holder

TotalEnergies est une compagnie multi-énergies mondiale de production et de fourniture d'énergies : pétrole et biocarburants, gaz naturel et gaz verts, renouvelables et électricité.

Context

La réduction des émissions de gaz à effet de serre est un challenge mondial dans lequel TotalEnergies est pleinement engagé. Nous avons mis en place différentes actions pour répondre à cet enjeu majeur dont le captage-stockage de CO2.

En effet, selon l'AIE, le captage-stockage de CO2 pourrait contribuer jusqu'à 14 % des réductions des émissions mondiales de gaz à effet de serre à horizon 2050, ce qui représente environ 120 milliards de tonnes de CO2 captées et stockées entre 2015 et 2050.

Lien : https://totalenergies.com/fr/dossiers/le-captage-stockage-de-co2-une-solution-prometteuse

Problem and goal

The ambition to sequester CO_2 in the sub-surface requires identification and delineation of aquifer units / **formations**. Pie-charts representing lithology and thickness of aquifer formations form the basis of maps which delineate an aquifer's spatial extension and quality.

In order to create such pie-charts, geoscientists consult analogue datasets – principally **final well reports / completion reports** & **composite well logs / completion logs** - which describe in text and images the well location (co-ordinates), lithologies and formation thickness.

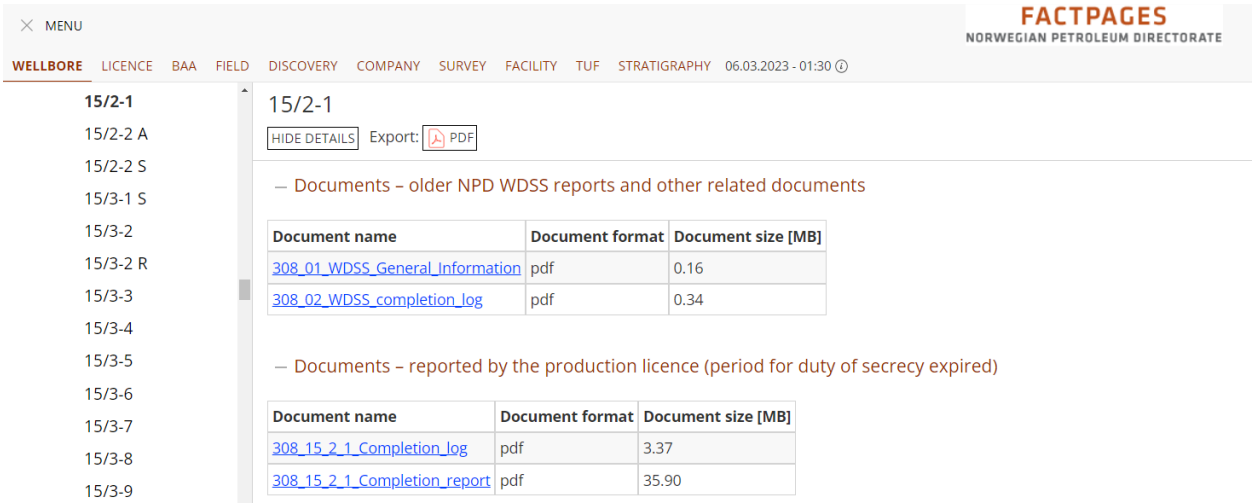
The reports are typically scanned .pdf images of typed reports. The text descriptions are qualitative and differ in the descriptive terms used; they are best used to understand dominant lithologies (sense check) but not to create quantitative pie-charts. The text descriptions of the formations also include reference to the formation **top** and base depths (the difference between the two depths corresponds to the measured formation thickness). The images in the reports / composite well logs are often of mediocre quality and noisy. As a result, the images would likely benefit from enhancement / de-noising prior to automated interpretation (example included where noise may be incorrectly interpreted as sandstone).

Once the image of the lithology column relating to the desired formation is identified, the image can be used to quantitatively assess the thickness of the different component lithologies and thus create a quantitative pie-chart. It is important that in the solution / workflow provided, the name of the formation can easily be changed (so that the workflow can produce results for different formations / account for different naming conventions).

As the aquifers considered for geological carbon storage are often extensive, a correspondingly large set of reports must be consulted at the outset of any study. It is therefore requested to automate the interrogation of these reports and extract quantitative lithological pie-charts (annotated with their measured thickness) which can be visualized on a map. This product will form the basis of an aquifer map and inform decisions regarding how appropriate a location is with regards to permanently storing CO_2.

Dataset

You can find all public wells on this url: https://factpages.npd.no/en

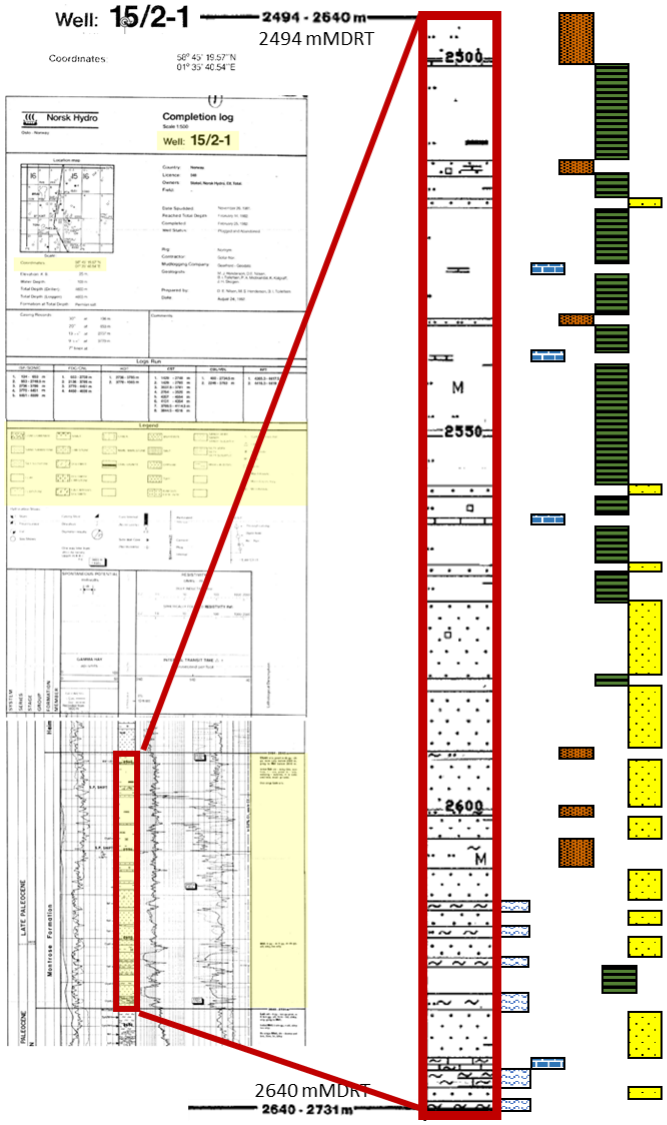


25 wells were also extracted to evaluate models (zipped folder: "*NO_Quad_15.7z*")

Below is an example of input and output:

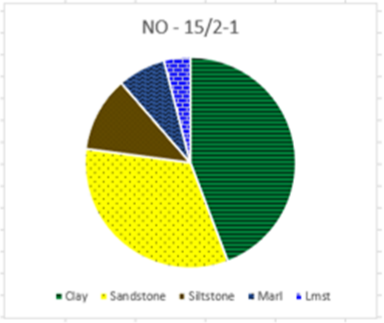
L'association Machine Learning
Data Science et Intelligence Artificielle

DATA BATTLE 2023



	Clay	Sandstone	Siltstone	Marl	Lmst	
Height	2.84	0.29	1.55	0.33	0.33	
	0.73	0.29	0.41	0.33	0.33	
	1.64	0.29	0.33	0.28	0.33	
	1.18	2.24	0.33	0.56	0.33	
	0.83	1.86	0.33	0.56		
	3.62	1.44	0.83	0.35		
	0.52	0.67				
	1.09	0.92				
	0.91	0.41				
	0.33	0.62				
	0.81	1.4				
		0.35				
Sub-Sum	15	11	4	2	1	
SUM			33			Sense Check
Litho %	44%	33%	12%	7%	4%	100%

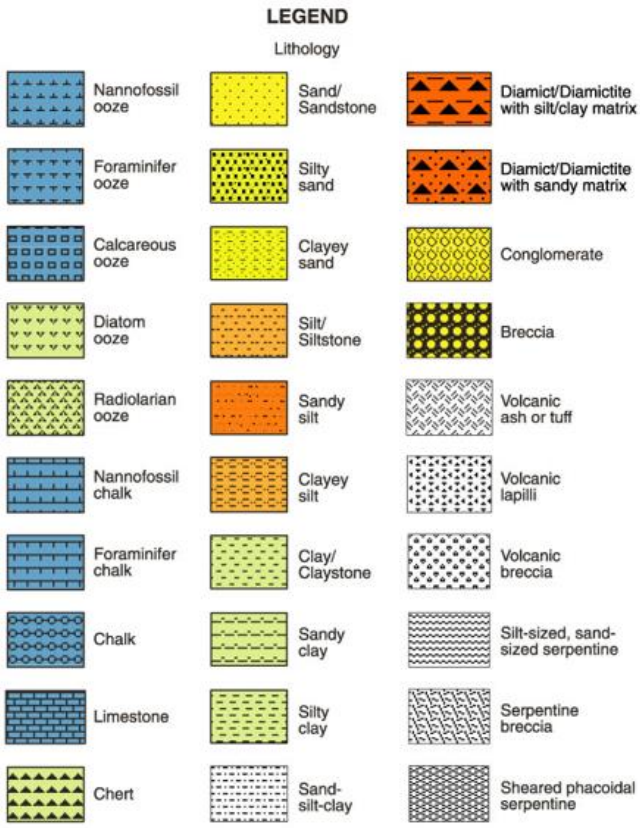
60 % 27 %



L'association Machine Learning
Data Science et Intelligence Artificielle

DATA BATTLE 2023

Pattern to be used to create the lithological pie-charts



Evaluation

Criteria to judge results:

1. Accuracy: +/- 10 % Sandstone
2. Accuracy: +/- 10 % Clay / Shale
3. Solution Efficiency: % of 10 blind wells + 1 live well during presentation
4. Data to which metrics above applied:
    - 25 Wells data provided (data provided + available on https://factpages.npd.no/en)

DATA BATTLE 2023

- 10 wells blind 1 hr before deadline (no data provided; https://factpages.npd.no/en)
- 1 live well during presentation (no data provided; https://factpages.npd.no/en)

"Reusability: based on Opensource" libraries

A notebook to test (a dashboard is plus).

Glossary

Geological Formation (Formation):

A body of rock with consistent physical characteristics (lithology) that distinguish it from adjacent bodies of rock, and which occupies a particular position in the layers of rock in a geographical region / sedimentary basin (stratigraphic column).

Formation Top (Tops):

A depth in a well (measured relative to a reference datum) at which geological formations are found in the subsurface.

Composite Well Log / Completion Logs:

All basic logs acquired in a well over different depth intervals spliced together to obtain single well log tracks per basic well log combined with lithological descriptions & other measurements (e.g., pressure, core etc).

Final Well Report / Completion Report:

Final report documenting all processes, observations, and results during the drilling of a well (including placing operations in chronological order).

Lithology:

The lithology of a rock unit is a description of its physical characteristics. Please note that traces elements / descriptions (e.g., pyrite) are not consequential to this challenge. The main lithologies of interest are:

Sandstone: Clastic sedimentary rock; composed of sand-sized (0.0625 to 2 mm) silicate grains.

Silt: Clastic sedimentary rock; A form of mud rock with a low clay mineral content, which can be distinguished from shale by a lack of fissility (tendency to split along flat planes of weakness).

Shale: Clastic sedimentary rock; composed of mud that is a mix of flakes of clay minerals and fragments of other minerals, especially quartz and calcite.

Clay: Clastic sedimentary rock; composed of clay-sized particles (less than 1/256 millimetre in diameter) and not laminated (layered) or easily split into thin layers (laminae / beds).

Marl: Clastic sedimentary rock; composed of a mixture of carbonate minerals, clays, and silt.

Limestone: Clastic sedimentary rock; composed of calcite and aragonite (different forms of calcium carbonate / $CaCO3$).

Dolomite: Clastic sedimentary rock; composed of calcite and aragonite with a high percentage of dolomite ($CaMg(CO3)2$).

Lithological Descriptions:

Bed: A coherent layer of sedimentary rock greater than 1 cm thick.

Laminae: A coherent layer of sedimentary rock less than 1 cm thick.

Stringer: A thin, discontinuous rock layer.