CREWES

www.crewes.org

A CREWES Data Science Initiative and the 2019 Projects

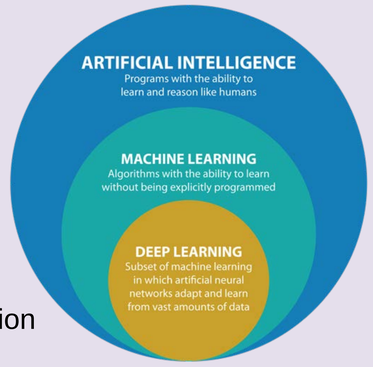
Marcelo Guarido*, Daniel Trad, and Kristopher Innanen
mguarido@ucalgary.ca

CREWES Data Science Initiative

What is the CREWES data science initiative?

The CREWES Data Science Initiative, or CREWES.AI, is the consortium participation on machine learning and data science world. The goal of the group is to deliver top-end research and applications of machine learning to the industry. The group will work together and in collaboration on different types of projects for seismic and other types of data from geoscience and oil & gas world, such as:

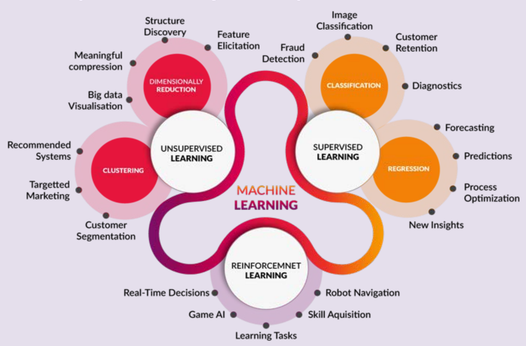
- Geophysical:
  - Seismic
  - Gravimetry
  - …
  - Multiphysics
- Petrophysical:
  - Wireline and LWD
  - Reservoir
- Engineering:
  - Production optimization
  - Drilling
- Earth Sciences:
  - Solid earth projects



The focus will be on mid term (5 to 12 months) and long term (1+ years) projects and can address the sponsors needs while keeping the academic freedom. Any CREWES student and staff can join the ML research group, and will be 100% focused on the project until its conclusion.

Mentorship programs to sponsor and short courses will also be available as the following:

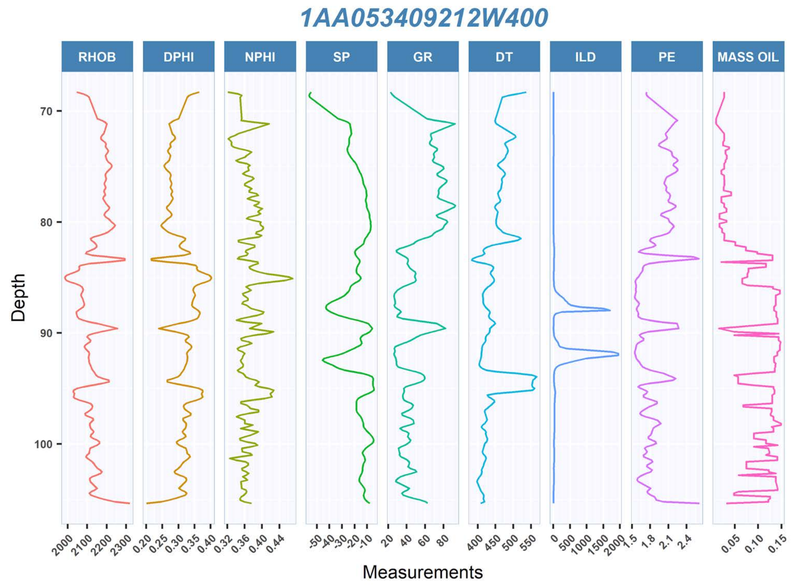
- Basics programming:
  - Python
  - R
  - MatLab
- ML programming:
  - General ML programming
- Custom ML programming:
  - Client specific ML programming
  - Dataset provided by the sponsor

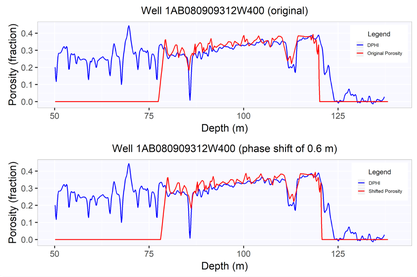


Machine Learning as a Tool to Predict the Mass of Oil from Well Logs

The Goal

Predict the fraction of mass of oil at each depth sample from the wireline logs. The 50 wells data used is from the athabasca Oil Field.



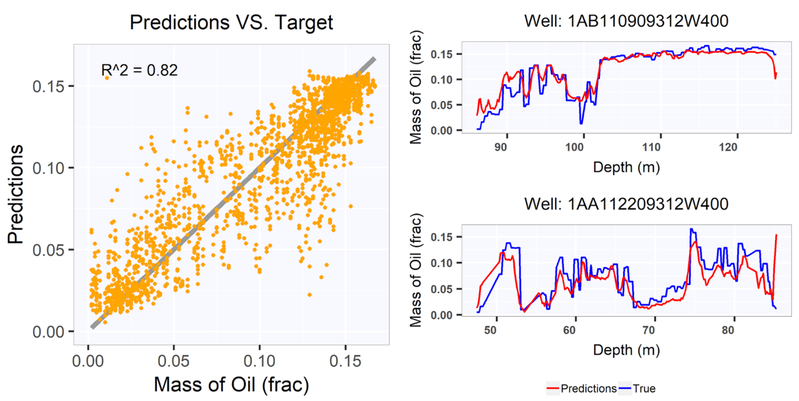


Data Preparation

Matching data from two different sources (in this case wireline logs and core analysis) assuming that the core porosity (red) and the density porosity (blue) are equivalent. The depth shift between the traces is calculated in several rolling windows and the final shift can be extracted from aa vector of possibilities. Each well is processed individually and each shift is independent.

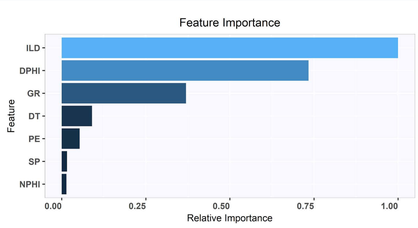
Modeling and Predictions

From the start, the features (well logs) presented non-linear relationship with the measured fraction of mass of oil from the cores. The gradient boosting regressor is a machine learning algorithm that can recognize such non-linear patterns and was used to predict the mass of oil. Predictions (red) over validation wells (blue) are robust, and the R^2 is 0.82.



Interpretation

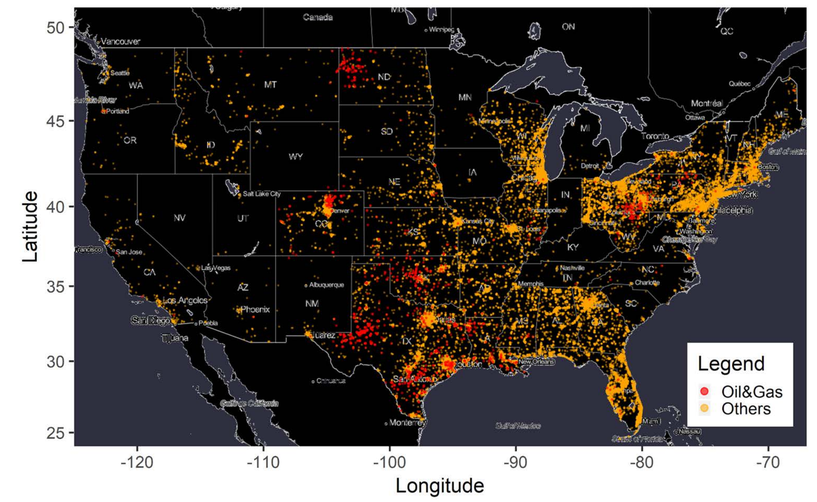
Predict the fraction of mass of oil using a tree based model (gradient boosting regressor), has the advantage to provide the importance of each feature. ILD and DPHI are pointed as the most important ones, while SP and NPHI here not important for the trained model.



Using Natural Language Processing and Machine Learning to Predict Severe Injuries Classification in the Oil and Gas Industry

The Goal

Classify the injury type in the oil & gas industry (red in the map) from the incident description by knowing the class in the other industries (orange in the map.

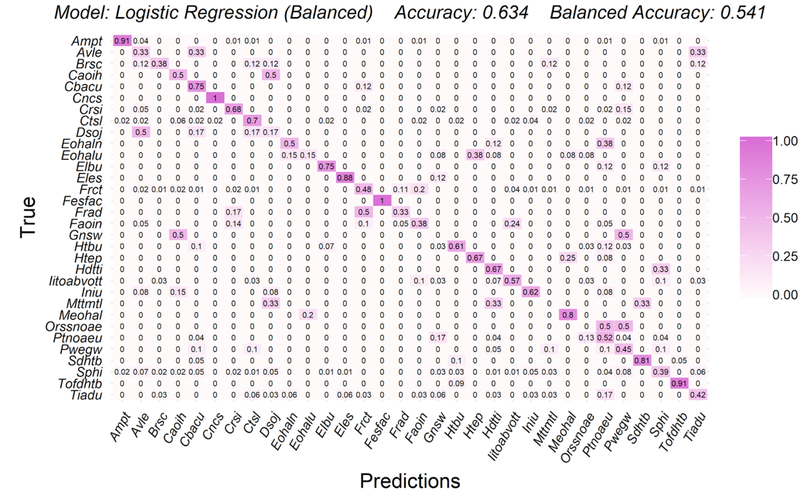


Modeling and Predictions

To convert descriptions to numerical variables, it was used the TF-IDF (term frequency–inverse document frequency), which weights each word inversely of its document frequency. All desirable important word is now a feature for modeling. At first, the TF-IDF weights were used for the predictions of the 32 classes and two models were tested: a *gradient booster classifier* and the *logistic regression*. The first has a better accuracy if disconsidering the imbalanced nature of the classes, meaning it classified better the most common classes, but failed for the least frequent ones. The logistic regression has a lower accuracy, but worked better in the overall scenario. The best prediction is reached using a logistic regression for balanced classes. and the confusion matrix below shows the accuracy for each class.

Confusion Matrix: Relative Values

Model: Logistic Regression (Balanced) Accuracy: 0.634 Balanced Accuracy: 0.541



Acknowledgments

The authors thank the sponsors of CREWES for continued support. This work was funded by CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 461179-13. We also thank GLJ Petroleum Consultants, specially Bill Spackman and Michael Morgan, for technical support and data acquisition. Finally, we thank Soane Mota dos Santos for all the knowledge share during very useful conversations.

NSERC CRSNG

UNIVERSITY OF CALGARY
FACULTY OF SCIENCE
Department of Geoscience