Predictive Analytics of Induced Seismicity from Saltwater Disposal Well Performance

Rhett Marcelo
University of Houston – Downtown
rhettmarcelo@acm.org

Kalyan Venugopal, Ph.D.
Subject Matter Expert
kalyan100673@gmail.com

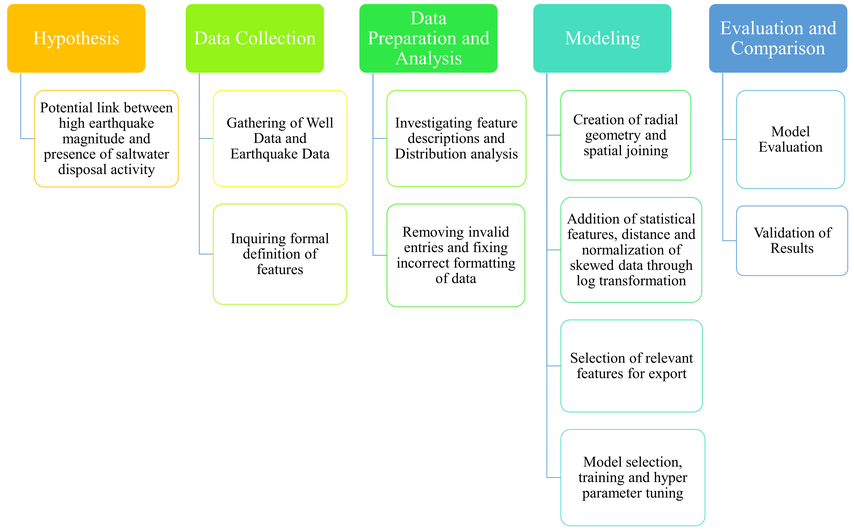
Dvijesh Shastri, Ph.D.
University of Houston – Downtown
shastrid@uhd.edu

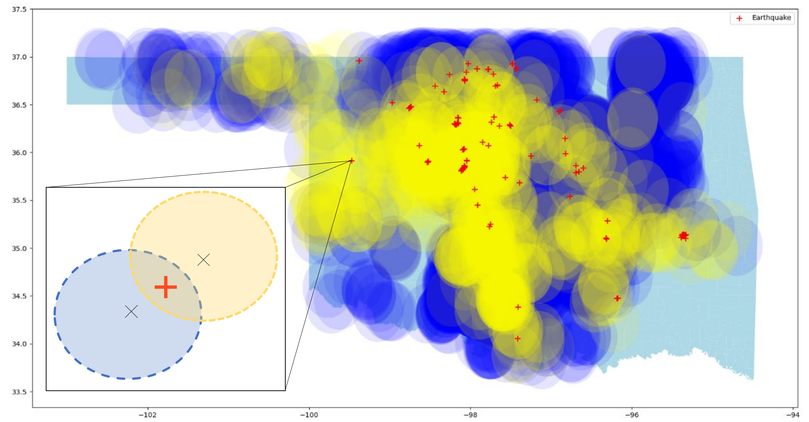
Abstract

The intention of this research project is to find a relationship between saltwater disposal wells and induced seismicity in the state of Oklahoma using a data-driven approach and methodology. Well and earthquake datasets are explored and analyzed, then merged through the application of spatial join method. Machine learning methods are built, tuned and used to predict the magnitude of an earthquake and its potential distance from each wellsite.

Introduction

Saltwater disposal wells (SWD) are specifically designed to inject large volumes of saltwater and produced water from the deep underground back into porous rock formations. Induced seismicity refers to earthquakes that are caused or triggered by human activity. In recent years, there has been growing concern about the link between SWDs and induced seismicity. The injection of large volumes of fluid into the subsurface can increase pore pressure in the surrounding rock formations, which can, in turn, reduce the frictional forces that hold the rocks together. This can lead to the reactivation of pre-existing faults or the creation of new ones, resulting in earthquakes.

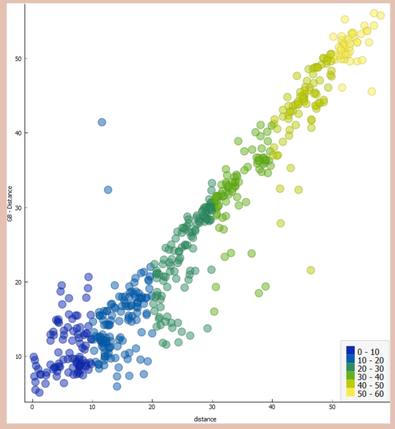




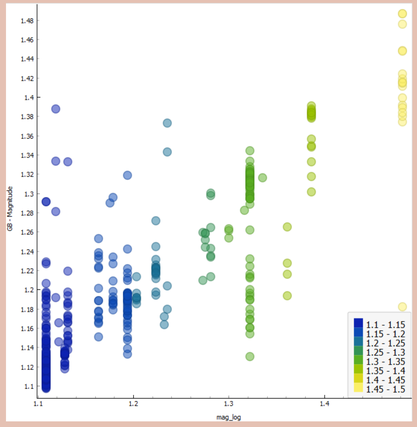
Spatial Joining of Well and Earthquake Lat/Lon in Oklahoma

Methodology

Saltwater disposal well data set is acquired from Oklahoma Corporation Commission and earthquake dataset is acquired from U.S. Geological Survey. Both datasets are cleaned, removing rows with invalid values and imputing empty spaces based on average and most frequent value. Using the GeoPandas library, both datasets are spatially joined based on arbitrary well distance radius. Statistical columns are added for normalization and feature engineering of skewed data such as log transformation and square root transformation. The preprocessing is done in Python and exported for predictive analytics. The data is loaded in Orange and setting distance and magnitude for different models as targets for Random Forest, Gradient Boosting and Multilinear Regression to learn and predict the magnitude and distance of the earthquake. Hyperparameter tuning is performed to increase the confidence of the learning and prediction.



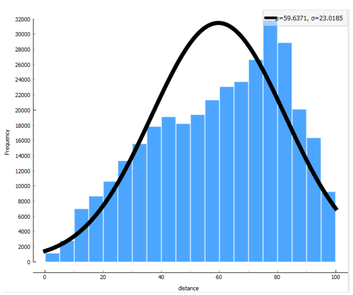
Distance Model



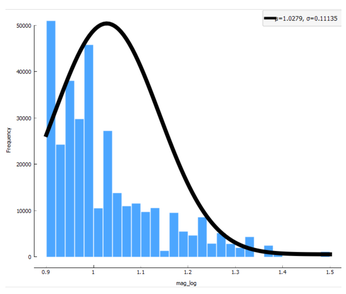
Magnitude Model

Regression Metrics

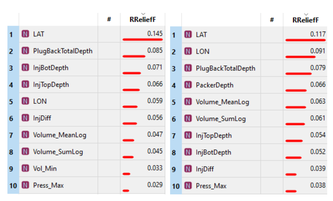
	Model	Train Time	Test Time	MSE	RMSE	MAE	R^2
Magnitude	Random Forest	1.226	0.073	0.00272	0.05219	0.02963	0.75450
	Multilinear Regression	0.123	0.020	0.00776	0.08813	0.07361	0.30000
	Gradient Boost	1.584	0.023	0.00225	0.04747	0.02587	0.79682
Distance	Random Forest	0.795	0.051	21.0762	4.59088	2.91728	0.910675
	Multilinear Regression	0.142	0.019	158.267	12.5804	10.1997	0.32923
	Gradient Boost	1.750	0.024	21.3113	4.61641	2.92943	0.909678



Distance Distribution



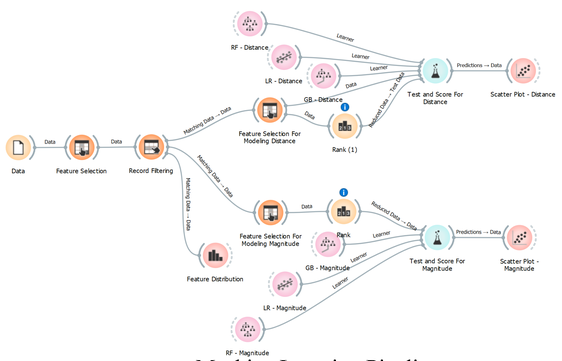
Magnitude Distribution



		RReliefF
1	LAT	0.145
2	PlugBackTotalDepth	0.085
3	InjBotDepth	0.071
4	InjTopDepth	0.066
5	LON	0.059
6	InjDiff	0.056
7	Volume_MeanLog	0.047
8	Volume_SumLog	0.045
9	Vol_Min	0.033
10	Press_Max	0.029

		RReliefF
1	LAT	0.117
2	LON	0.091
3	PlugBackTotalDepth	0.079
4	PackerDepth	0.066
5	Volume_MeanLog	0.063
6	Volume_SumLog	0.061
7	InjTopDepth	0.054
8	InjBotDepth	0.052
9	InjDiff	0.039
10	Press_Max	0.038

RReliefF Feature Ranking



Machine Learning Pipeline

Conclusion and Future Work

Machine learning models that can predict seismicity and radial distance with confidence can infer an existing relationship between saltwater disposal wells and induced seismicity. Although complex, non-linear and context-specific, a data-driven approach suggests that there is evidence that the operation of saltwater wells can potentially cause earthquakes in some cases. The future work involves extending the scope for different years and states.

Acknowledgements

This material is based upon work supported by the NSF Grant IIS-2123247. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.