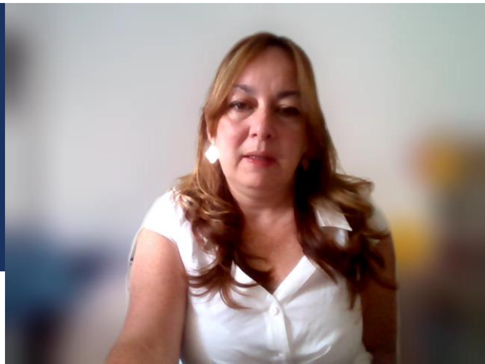
UTAH FORGE
U.S. Department of Energy

SPE International
Gulf Coast Section
Data Analytics Study Group

S-Team

PIVOT 2022 SPE Datathon
22nd July 2022



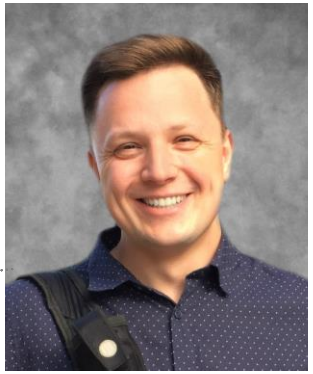
Who we are

S-team is a multidisciplinary team located in different worldwide locations

NABEEL MUHAMMADY
Geophysics

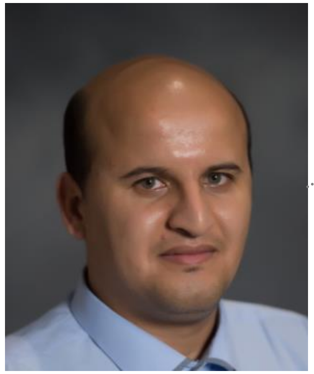
JERJES PORLLES
Petroleum Engineer

ALEXEY BORISENKO
Petroleum Engineer



SONIA LOPEZ KOVACS
Reservoir Engineer

OMAR BAKELLI
Petroleum Engineering



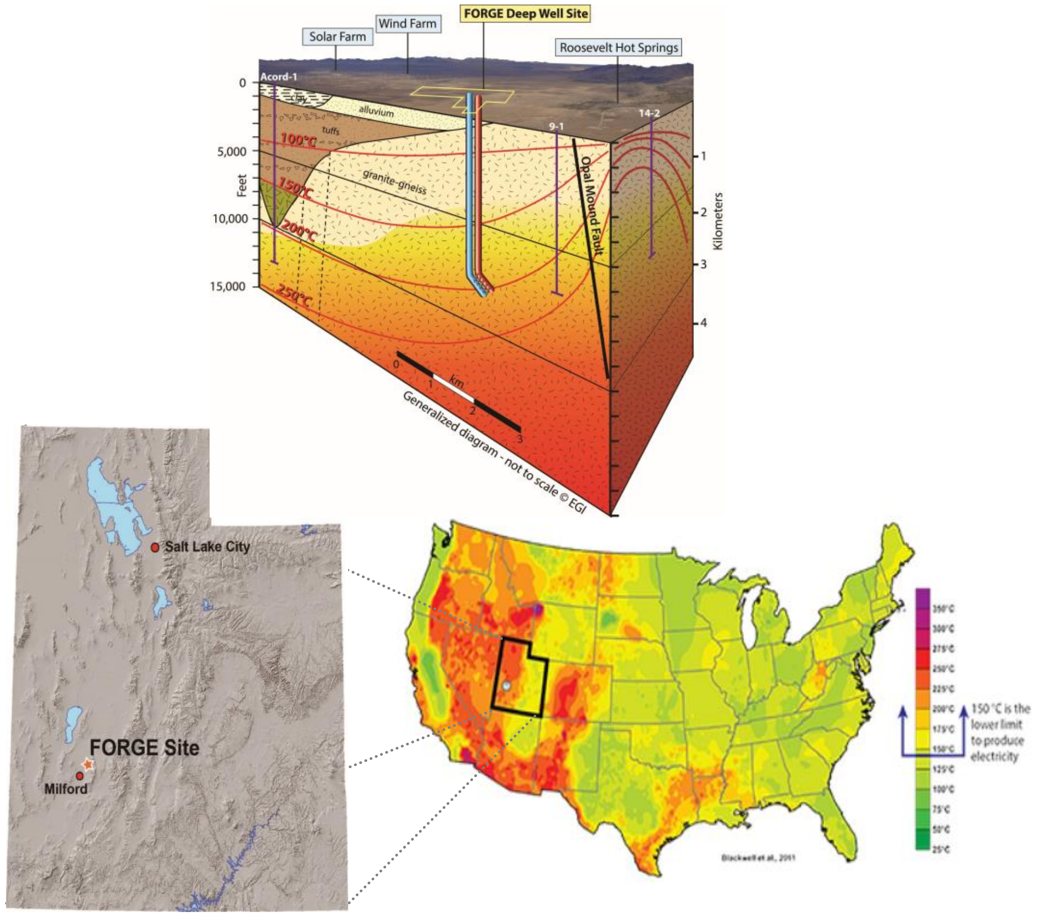
MOISES VELASCO
Petroleum Engineer
Ph.D. student

JESSE NYOKABI
Engineering

Problem statement

Utah Forge

The Utah FORGE site is located in south-central Utah within the Utah Renewable Energy Corridor.

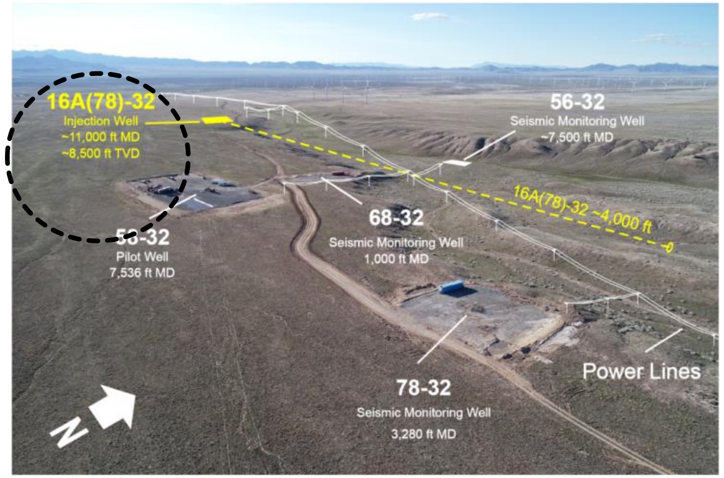


Source: https://utahforge.com

Evaluate a hypothetical scenario where the FORGE location is an EGS production field, and the first FORGE highly deviated well (16A) is the injection well.

Objective

Find the optimum placement of the production well that **maximizes the net energy and electrical-power** output over a 20-year project lifespan



Utah FORGE project site infrastructure, drone shot (source: Utah FORGE)

Workflow

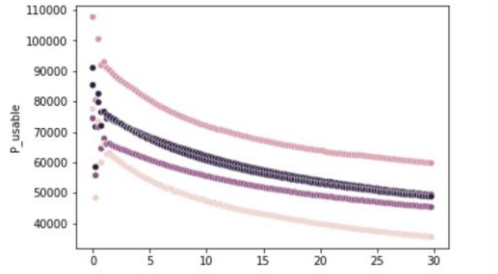
DATA ACQUISITION → **DATA PREPARATION** → **MODEL TRAINING** → **RESULTS**

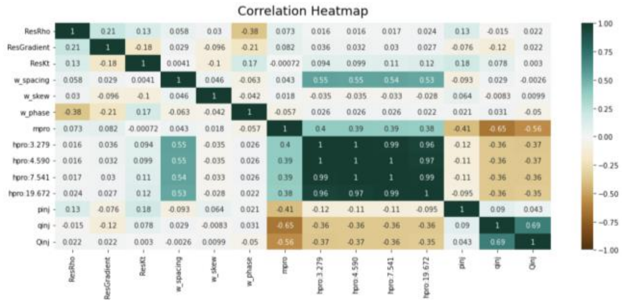
Inputs results FORGE

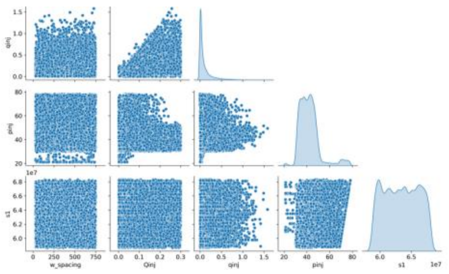
FILE / GROUPS

- Inputs results main 30 years
- Inputs results medium 20 years
- Inputs results short1year

METHODOLOGY







OBJECTIVE FUNCTION : Enthalpy Cumulation

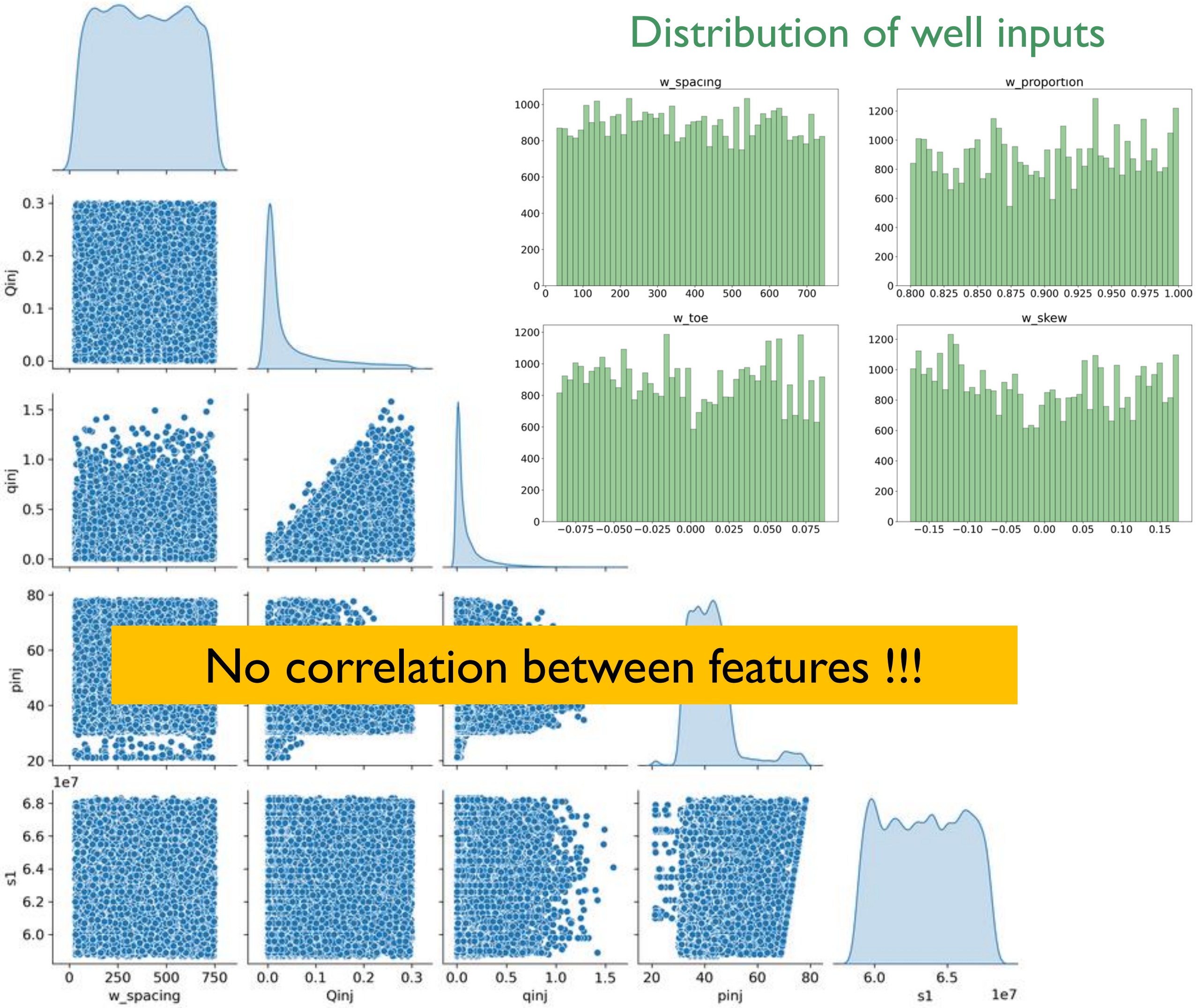
Find:

- ✓ Optimum well placement.
- ✓ Well placement to avoid.

Exploratory Analysis

Bivariate analysis

Distribution of well inputs



No correlation between features !!!

Data preparation

- Imputation (KNN)
- NaN values
- Outliers

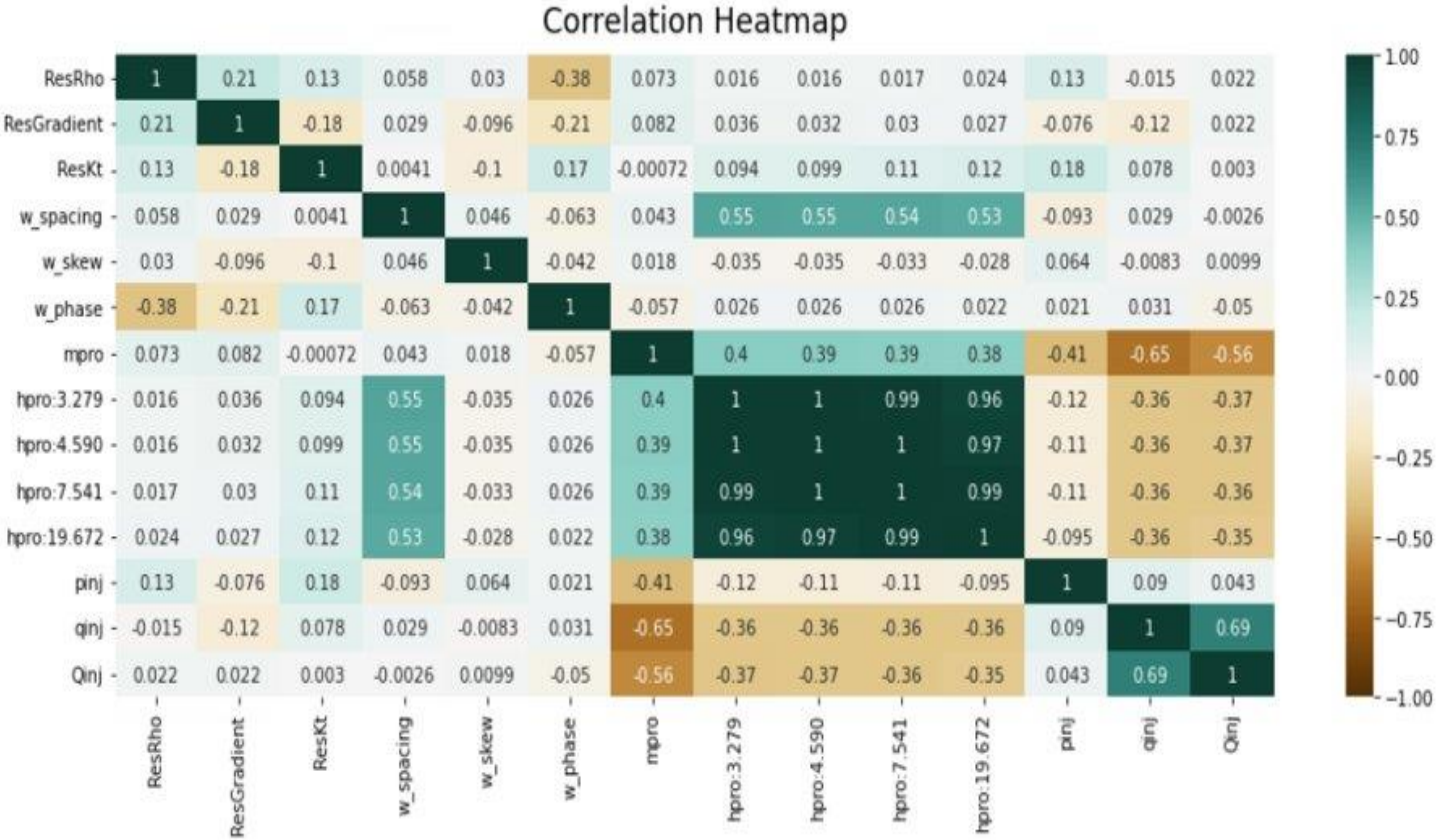
Feature selection

- Partial correlation
- Mutual information
- PCA (dimensionality reduction)

Feature engineering

- Equations
- Physics in geothermal production

Correlation Matrix



Objective functions

In this competition, we evaluate a hypothetical scenario where the FORGE location is an EGS production field and the first FORGE highly deviated well (16A) is the injection well. Based on the provided model results, we seek the theoretical optimum placement of the production well that maximizes the likelihood of achieving maximum net-energy and electrical-power output over a 20 year project lifespan, accounting for parasitic losses.

- Timeseries without losses

$$P_{usable} = 0.13(dhout + (h_{95C} - h_{120C})mpro)$$

- Cumulative without Losses

$$P_{usable} = \frac{1}{20yr}\int_0^{20} 0.13(dhout + (h_{95C} - h_{120C})mpro)dt$$

- Timeseries with Losses

$$P_{usable} = 0.13(dhout + (h_{95C} - h_{120C} - v5(Pinj - Pwhp))mpro)$$

where:

P_{usable} - amount of enthalpy, kJ/s

$dhout$ - enthalpy-rate change from injection to production, kJ/s

h_{95C}, h_{120C} - reference enthalpies (calculated), kJ/kg

$v5$ - specific volume (thermodynamic term), m^3/kg

P_{inj} - pressure of injected fluid, MPa (x10^3 kJ/m^3)

P_{whp} - power plant inlet pressure, MPa (x10^3 kJ/m^3)

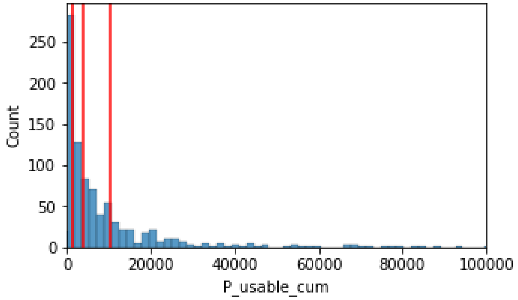
$mpro$ - production mass flowrate, kg/s

Classification problem solving

Objective function selection

$$P_{usable} = \frac{1}{20yr} \int_0^{20} 0.13(dhout + (h_{95C} - h_{120C})mpro)dt$$

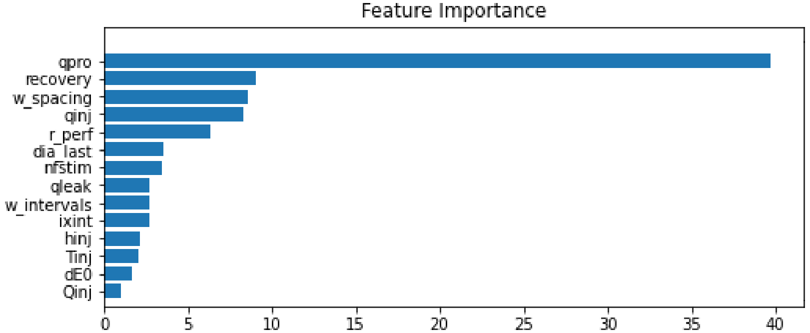
Splitting dataset by category based on P_usable calculated



Solving classification problem with ML method based on gradient boosting on decision trees algorithm – **f1-score: 0.89**

CatBoost

Highlighting important features based on ML model output results



The most important parameters for optimal scenario over 20-year project lifespan (P_usable > 10500 kJ/s)

	qpro	recovery	w_spacing	qinj	r_perf	dia_last	nfstim	qleak	w_intervals	ixint	hinj	Tinj	dE0	Qinj
count	221.000000	221.000000	221.000000	221.000000	221.000000	221.000000	221.000000	221.000000	221.000000	221.000000	221.000000	221.000000	2.210000e+02	221.000000
mean	-0.155352	-1.010498	612.647964	0.171984	122.517647	4532.904977	2.873303	-0.036421	4.135747	4.262443	289.425339	59.949321	1.762127e+06	0.081488
std	0.099026	0.724619	261.531804	0.144571	52.290859	6390.773496	2.572982	0.062100	1.286163	1.382978	75.933943	18.744887	3.072510e+05	0.069692
min	-0.660000	-4.780000	56.200000	0.000000	11.200000	290.000000	0.000000	-0.360000	1.000000	1.000000	150.000000	26.400000	1.190000e+06	0.000541
25%	-0.189000	-1.000000	424.000000	0.091400	84.900000	1470.000000	1.000000	-0.045200	3.000000	3.000000	233.000000	46.100000	1.460000e+06	0.030500
50%	-0.128000	-0.938000	630.000000	0.128000	126.000000	2260.000000	2.000000	-0.010500	4.000000	4.000000	267.000000	56.300000	1.840000e+06	0.055200
75%	-0.092400	-0.750000	847.000000	0.195000	169.000000	3530.000000	5.000000	-0.000250	5.000000	5.000000	365.000000	79.700000	2.030000e+06	0.117000
max	-0.039300	1.000000	996.000000	0.980000	199.000000	31700.000000	10.000000	0.000000	6.000000	7.000000	412.000000	85.000000	2.600000e+06	0.295000

Well placement (optimum)

	pin	w_spacing	w_length	w_azimuth	w_dip	w_skew	w_count	w_toe	w_proportion	w_phase	w_intervals
139	110998313	443.0	1110.0	1.83	0.438	-0.1610	1.0	0.0125	0.998	3.14	5.0
263	249248828	769.0	1110.0	1.83	0.438	0.0747	1.0	0.0458	1.030	3.14	5.0

The most important parameters for not optimal scenario over 20-year project lifespan (P_usable < 1300 kJ/s)

	qpro	recovery	w_spacing	qinj	r_perf	dia_last	nfstim	qleak	w_intervals	ixint	hinj	Tinj	dE0	Qinj
count	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	216.000000	2.160000e+02	216.000000
mean	-0.006508	-0.842690	454.313426	0.010464	90.841296	1839.828704	1.009259	-0.003986	3.222222	3.231481	286.518519	60.416667	1.749676e+06	0.005813
std	0.018690	0.222428	291.400142	0.039345	58.256684	1261.234354	1.346799	0.023088	1.480624	1.510315	82.141396	20.030382	3.051151e+05	0.020583
min	-0.202000	-1.370000	33.400000	0.000586	6.680000	173.000000	0.000000	-0.241000	1.000000	1.000000	149.000000	26.400000	1.190000e+06	0.000501
25%	-0.004605	-1.000000	232.000000	0.002508	46.375000	780.500000	0.000000	-0.001150	2.000000	2.000000	203.000000	40.600000	1.510000e+06	0.000705
50%	-0.003250	-1.000000	408.000000	0.003770	81.500000	1510.000000	1.000000	0.000000	3.000000	3.000000	300.000000	61.700000	1.730000e+06	0.001075
75%	-0.002033	-0.748750	668.000000	0.005595	133.500000	3080.000000	1.000000	0.000000	4.000000	4.000000	370.000000	81.900000	2.000000e+06	0.002140
max	-0.000587	-0.218000	996.000000	0.376000	199.000000	6740.000000	9.000000	0.000000	6.000000	7.000000	394.000000	87.200000	2.600000e+06	0.202000

Well placement (to avoid)

	pin	w_spacing	w_length	w_azimuth	w_dip	w_skew	w_count	w_toe	w_proportion	w_phase	w_intervals
84	466876610	515.0	1110.0	1.83	0.438	-0.084600	1.0	-0.03950	0.882	0.0	2.0
778	393261445	381.0	1110.0	1.83	0.438	0.000151	1.0	-0.00245	1.080	0.0	3.0

Conclusions and Recommendations

- ✓ An important part of the machine learning process is the analysis of the data. We could not see an evident correlation due to the complex distribution of the properties. We evaluated the data to find missing information and outliers.

- ✓ Our team developed a procedure to determine a theoretical optimum well placement of the production well and the well placement parameters to avoid. We selected the cumulative amount of enthalpy as the objective function. Many objective functions were shown, which were related to an amount of enthalpy calculation.

- ✓ In order to find what parameters are the most important machine learning method was used. This method is based on gradient boosting on the decision trees algorithm – CatBoost.

- ✓ We got 14 parameters for optimal and not optimal scenarios over a 20-year project lifespan.

- ✓ What did we learn?
  - ✓ How we can apply data science in geothermal energy, having a real case (Utah-FORGE).
  - ✓ Work as a team, recognize our skills, and organize our schedules (Different time zones)

References

1. https://utahforge.com/outreach/education/geothermal-resources-lecture-series/
2. https://publications.mygeoenergynow.org/grc/1033913.pdf
3. https://utahforge.com/2018/06/14/u-gets-140m-grant-for-geothermal-research/

Thanks