NOVA IMS

Information Management School

Advanced Topics in Geospatial Analysis

Acreditações e Certificações da NOVA IMS

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA IMS Information Management School

Case Study: Investigation of Geothermal Potential with Machine Learning in Mainland Portugal

Supervisors:

Roberto Henriques

Vicente Tang

Filiberto Pla

Matheus Nascimento

matheus.lpd.nascimento@gmail.com

19/05/2022

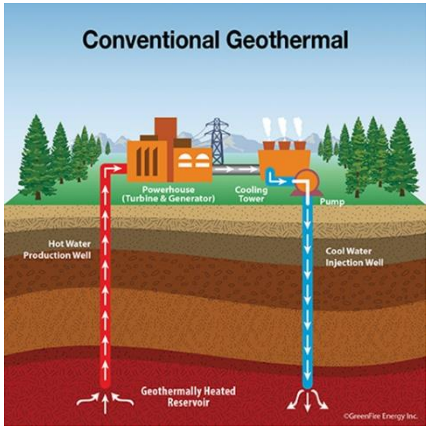
Acreditações e Certificações da NOVA IMS

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

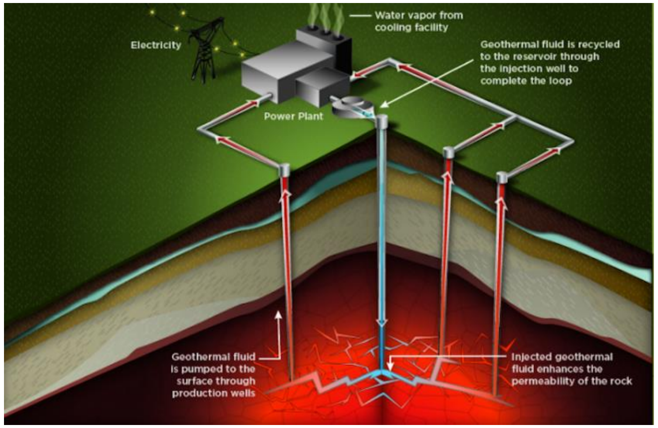
Presentation Structure

Introduction

- **Geothermal resources:**
  - **Heat stored inside the Earth's crust**
    - **Hot ground water**
    - **Heat in subsurface layers (EGS)**
  - **Used to generate electricity**
  - **Used directly**
    - **Heating of buildings and industrial processes**
  - **Renewable, cleaner and relatively constant**



www.greenfireenergy.com



www.thinkgeoenergy.com

Introduction

- ## Geothermal exploration:

  - **Search for places with hot water or heat that can be economically explored**

  - **Requires many resources**

  - **Traditionally achieved through:**
    - **Field surveys and drilling campaigns**
    - **Geophysical, geological and geochemical analysis**

www.thinkgeoenergy.com

Introduction

- **Methods to locate promising areas and reduce risks:**
  - **GIS-based spatial analysis**
    - **Geo statistics**
    - **GIS-based Multi criteria decision analysis**
  - **Machine Learning models**

Introduction

- **Study area: Mainland Portugal**

  - **Continental land with 89 015 km^2**
  - **Low enthalpy systems explored for direct use**

  - **Recent efforts to: improve knowledge, foment future geothermal explorations.**
    - **Geothermal Atlas of Portugal (Ramalho & Correia, 2015)**
    - **Promising theoretical EGS potential in the Iberian Peninsula (Chamorro et al., 2014)**

Introduction

- This study couples GIS and ML to assess the potential for geothermal exploration by predicting:

  - **Geothermal gradient (°C/km)**
  - **Surface heat flow density (mW/m^2)**

Introduction

- ## Goal and objectives:

  - **Evaluate if labels can be predicted using publicly available data on geological, hydrogeological, geophysical, terrain, and weather information by ML models**
  - **Compare the results of two different algorithms**
  - **Identify promising features to predict labels and interpretate possible relationships among them**

Literature Review

- **Geothermal Potential:**
  - **Estimated calculating important physical parameters of the ground and subsurface**
  - **Using different weighted criteria data to rank different areas regarding their potentiality**

(Shahdi et all., 2021)

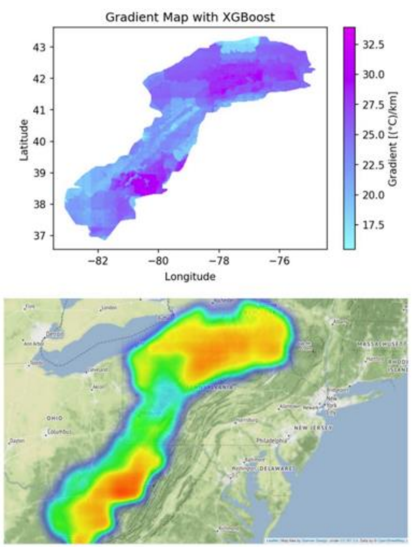
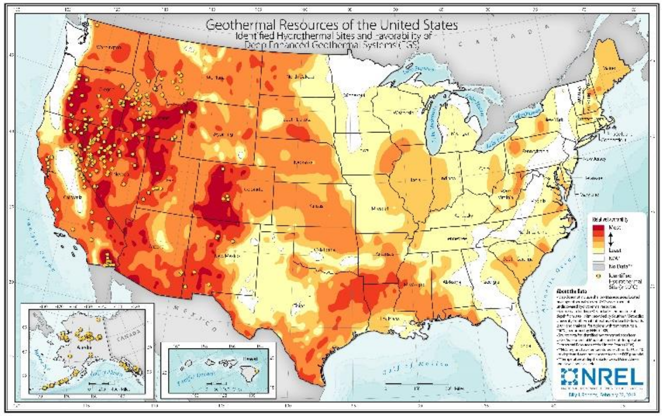


Fig. 12 Geothermal gradient map using XGBoost model. The gradient has the unit of $\frac{°C}{km}$.



www.nrel.gov

Literature Review

- **GIS-based techniques:**

  - **Studies successfully estimated suitability using MCDA:**
    - **AHP (Tinti et al., 2018, Omwenga, 2020)**
    - **Weighted overlay (Trumpy et al., 2015, Fahil et al., 2020)**

    - **Managed to identify or confirm important criteria data linked to geothermal potential**

    - **However, they are not used to estimate physical parameters values. ML is being used for this purpose**

Literature Review

- **Machine Learning:**

  - **Provided promising results in analysis of renewable energies exploration and suitability (Shahab & Singh, 2019, Lai et al., 2020)**

  - **In the Geothermal field:**
    - **Predict important thermal properties (Arola et al., 2019; Assouline et al., 2019; Gangwani et al., 2020; Mohamed et al., 2015; Shahdi et al., 2021)**
    - **Classify suitability indexes (Coro & Trumpy, 2020)**

Literature Review

- Machine Learning algorithms:

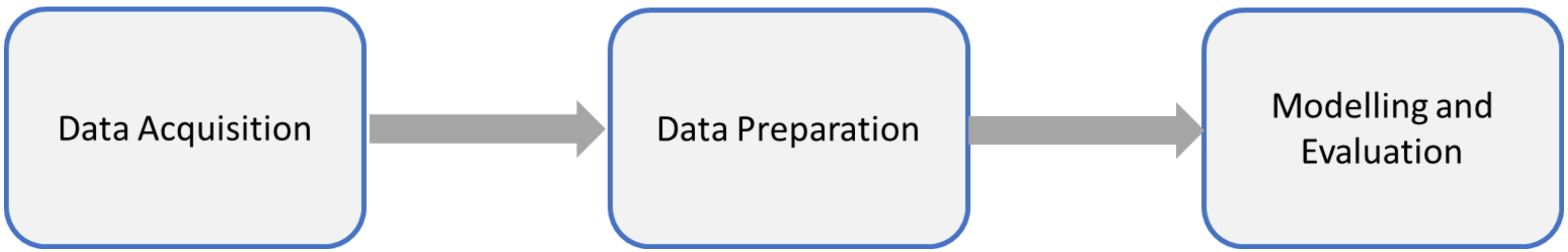
  - Random Forests (RF)
  - Extreme Gradient Boosting (XGB)
  - Gradient Boosted Regression Tree
  - Maximun Entropy Modeling
  - Artificial Neural Networks
  - Ridge Regression

Literature Review

- **Geothermal Related Variables/Criteria:**

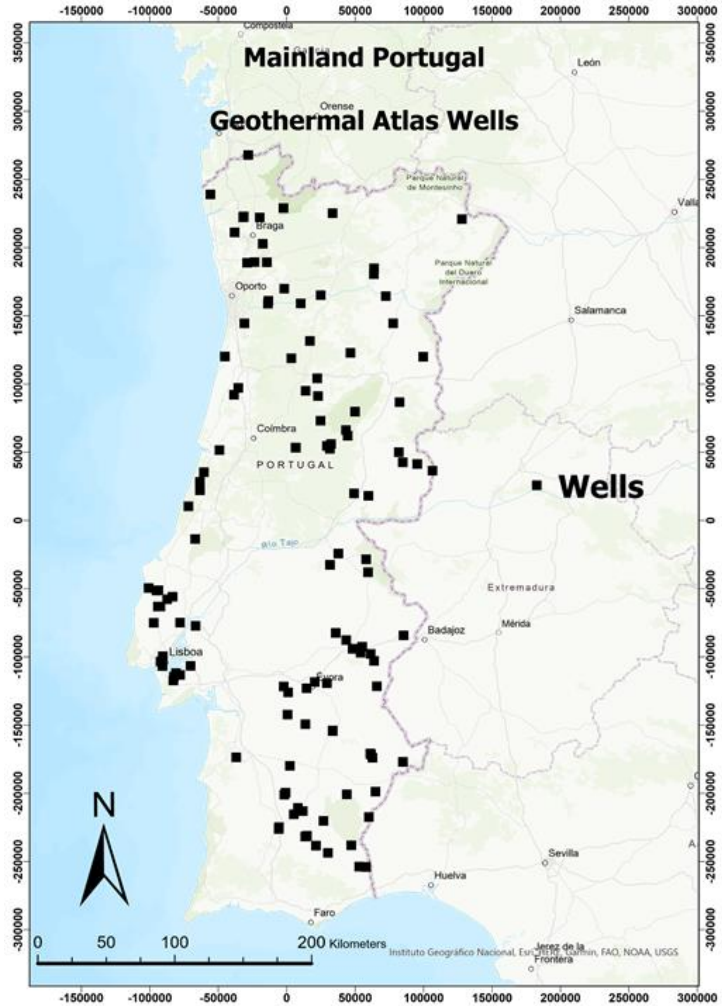
  - **Lithological and structural setting**
  - **Drainage density**
  - **Soil texture**
  - **Permeability**
  - **Bouguer gravity anomaly**
  - **Surface air temperature**
  - **Seismic activity**

Methodology



Methodology

- **Collected Data:**
  - **Labels were provided by LNEG and from the Geothermal Atlas of Portugal**
  - **138 points from different wells where geothermal gradient and surface heat flow density were measured**



Methodology

- ## Collected Data:

Collected Data	Source	Format
Geothermal wells	LNEG, Geothermal Atlas of Portugal	Shapefile
Elevation (DEM)	SRTM	Raster
Bouguer Gravity Anomaly	IRENA	Raster
Air Temperature 2m	Global Solar Atlas	Raster
Geological Map	LNEG	Shapefile
Faults Map	LNEG	Shapefile
Quaternary Faults	Quaternary Faults Database of Iberia	Shapefile
Soil Parental Material	University of Lisbon`s Superior Institute of Agronomy (ISA)	Shapefile
Soil Texture	University of Lisbon`s Superior Institute of Agronomy (ISA) EPIC WEBGIS	Raster
Potential Permeability	University of Lisbon`s Superior Institute of Agronomy (ISA) EPIC WEBGIS	Raster
Hydrogeological Units	National System of Environment (SNIAmb).	Shapefile
Seimic Intensity Zones	National System of Environment (SNIAmb).	Shapefile

Methodology

- ## Data preparation:

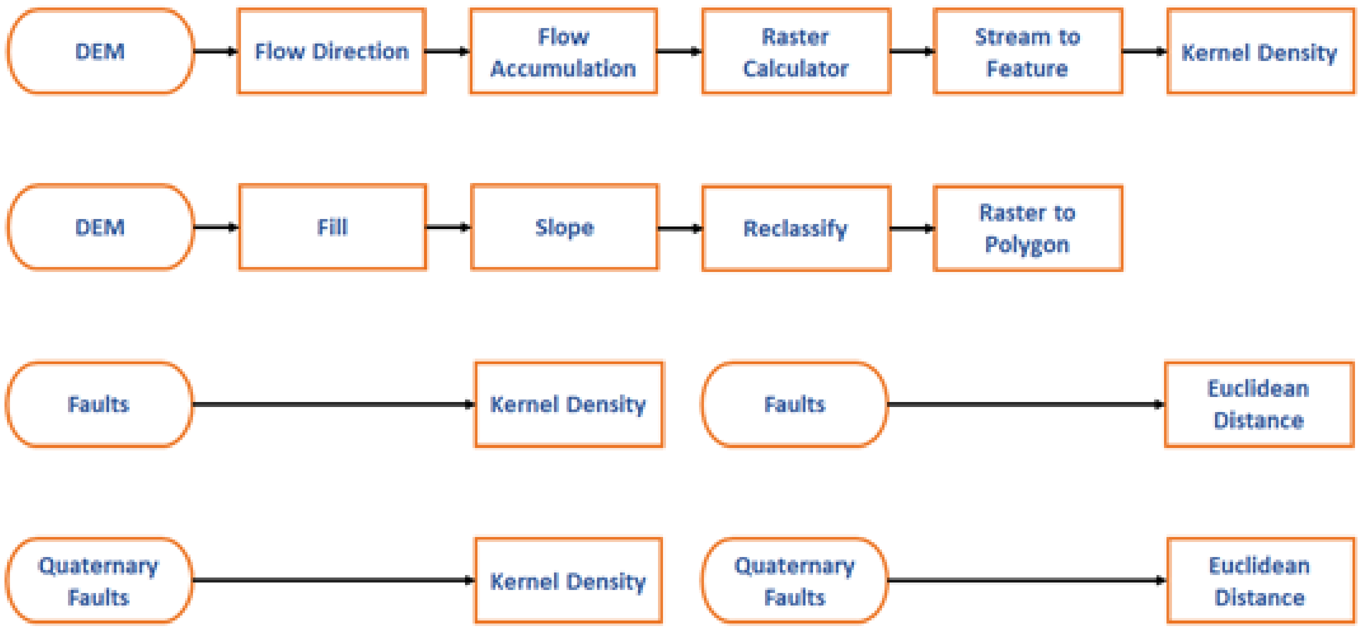
  - **Projected to ETRS 1989 Portugal TM06 coordinate system**
  - **Clipped to cover Mainland Portugal**

  - **Aggregated in two vector grids:**
    - **5x5 km spatial resolution**
    - **2,5x2,5 km spatial resolution**

  - **Grids created resampling the DEM. Elevation, latitude and longitude of the centroids were added as features.**

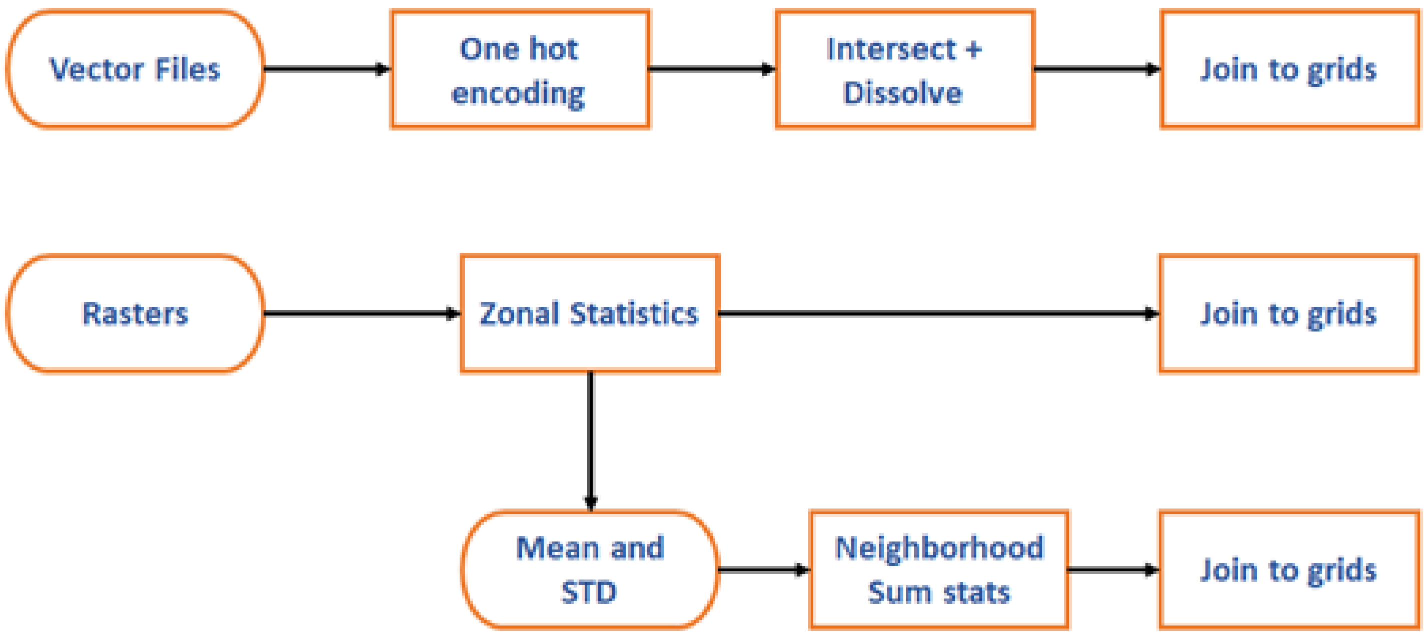
Methodology

- **Data preparation:**



Methodology

- ## Data preparation
  - **Categorical raster converted in Shapefiles**
  - **Raster with very low spatial resolution resampled**
  - **Missing values filled with nearest neighbors**



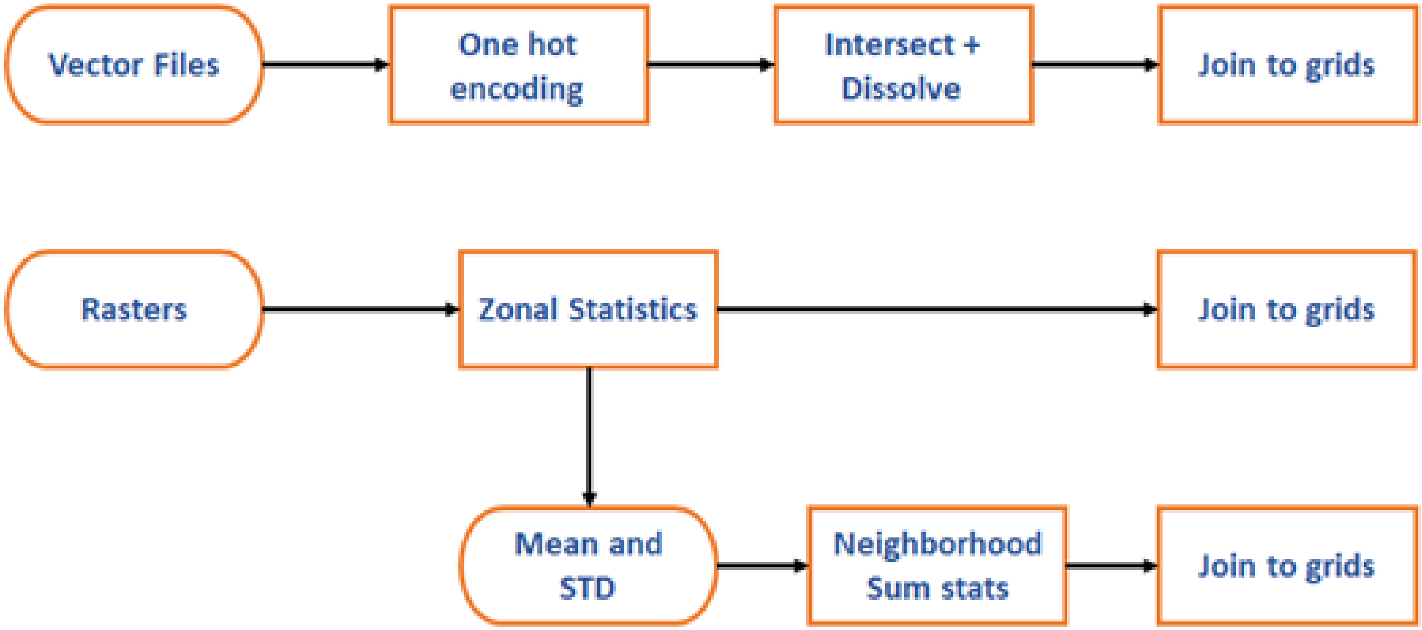
Methodology

- ## Data preparation
  - **One hot encoding**

Soil_Alluvia	Soil_Basic_heavy	Soil_Basic_very_heavy	Soil_Intermediate_loamy
0	0	0	0
1	0	1	1
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
1	0	0	0
0	1	0	0
0	0	0	0

Methodology

- ## Data preparation
  - **Categorical raster converted in Shapefiles**
  - **Raster with very low spatial resolution resampled**
  - **Missing values filled with nearest neighbors**



Methodology

- **Labels were summarized within the grids and outliers were removed following the 3*sigma rule**

Geothermal Gradient (°C/km)

All			Outliers removed		
Grids:	2.5x2.5 km	5x5 km	Grids:	2.5x2.5 km	5x5 km
count	**117**	**109**	**count**	**114**	**107**
mean	21,977	22,003	mean	21,406	21,573
std	**6,364**	**6,12**	**std**	**5,338**	**5,275**
min	10	10,448	min	10	10,448
max	48	48	max	37	37
25%	17,38	17,383	25%	17,245	17,382
50%	21	21	50%	20,8	20,6
75%	26	24,7	75%	24,467	24,512

Surface Heat Flow Density (mW/m²)

All			Outliers removed		
Grids:	2.5x2.5 km	5x5 km	Grids:	2.5x2.5 km	5x5 km
count	**117**	**109**	**count**	**115**	**108**
mean	63,852	63,42	mean	62,15	62,305
std	**22,673**	**21,446**	**std**	**18,538**	**18,095**
min	25	25	min	25	25
max	183,84	183,84	max	114,863	116,65
25%	50,55	50,55	25%	50,333	50,441
50%	60,842	60,99	50%	60	60,916
75%	73,578	72,5	75%	73,306	72,304

Methodology

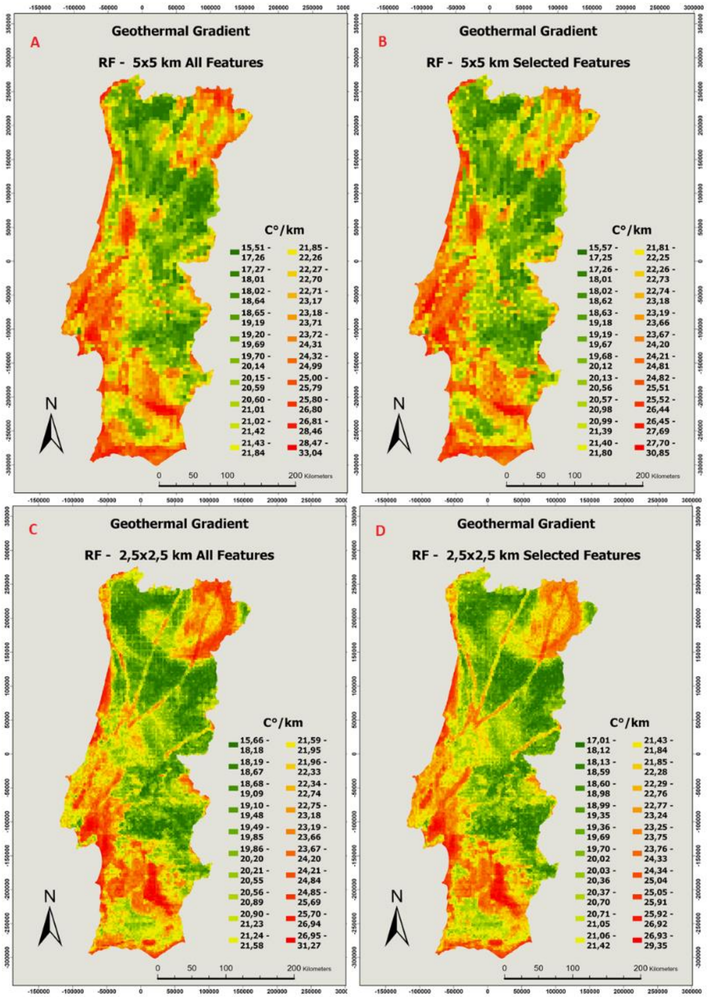
- RF and XGB Regressor were utilized to predict both labels
- Scikit Learn Python Module and Xgboost library

- Modelling and evaluation:
    - 5-fold Cross Validation
    - R^2, RMSE, and MAE
    - Grid Search CV – find best hyperparameters
    - Feature Selection with Features importance and Select from Model
    - Permutation Importance

Results

- ## Random Forests Models:

Geothermal Gradient				
	Random Forests (5x5 grid)		Random Forests (2.5x2.5 grid)	
Hyperparameters	All Features	Selected from Model	All Features	Selected from Model
Random State	3	2	2	2
Criterion	Squared error	Squared error	Absolute error	Absolute error
N Estimators	100	250	100	100
Max Depth	5	5	5	5
Min Samples Leaf	1	2	1	2
Min Samples Split	3	2	2	2
Errors				
MAE	3,746717631	3,66887079	4,058596306	4,016927566
RMSE	4,7220471	4,706783957	5,151928474	5,137863459
R2	0,201384987	0,215191652	0,039130307	0,045124678



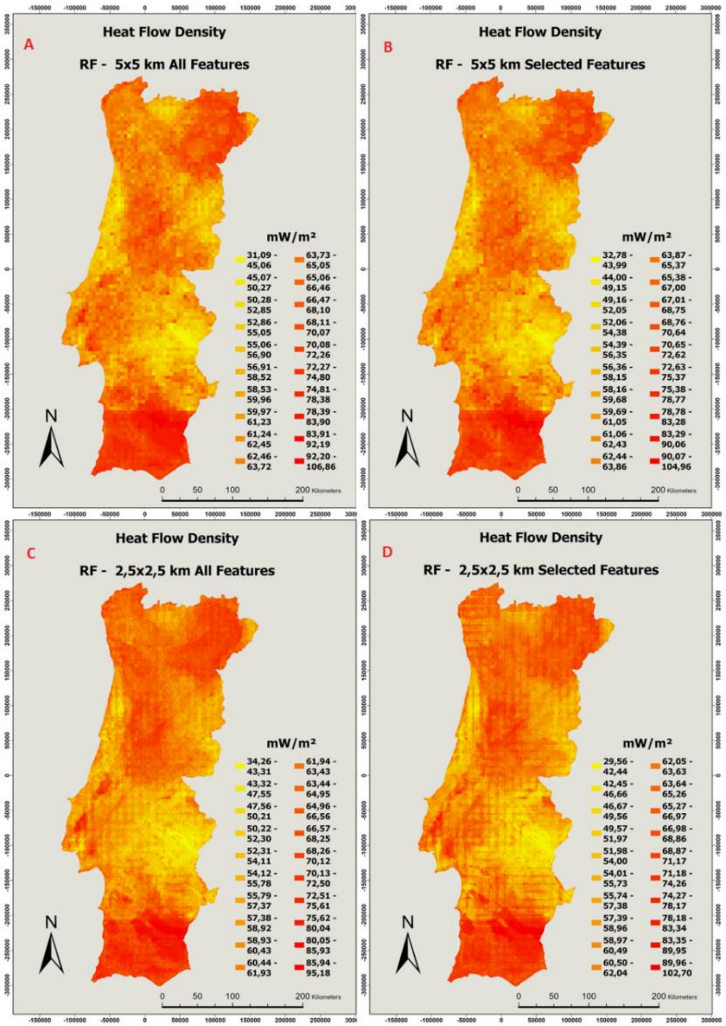
- **Values range: 15,57 – 30,85 °C/km**

Results

- ## Random Forests Models:

Heat Flow Density				
	Random Forests (5x5 grid)		Random Forests (2.5x2.5 grid)	
Hyperparameters	All Features	Selected from Model	All Features	Selected from Model
Random State	4	2	4	3
Criterion	Absolute error	Absolute error	Absolute error	Squared error
N Estimators	250	200	100	100
Max Depth	10	15	15	10
Min Samples Leaf	1	1	2	1
Min Samples Split	2	3	2	3
Errors				
MAE	12,38395917	12,0677001	13,06006468	12,92838859
RMSE	15,33868577	15,03292005	16,13293413	16,09533897
R2	0,221323496	0,248328421	0,223639738	0,231819702

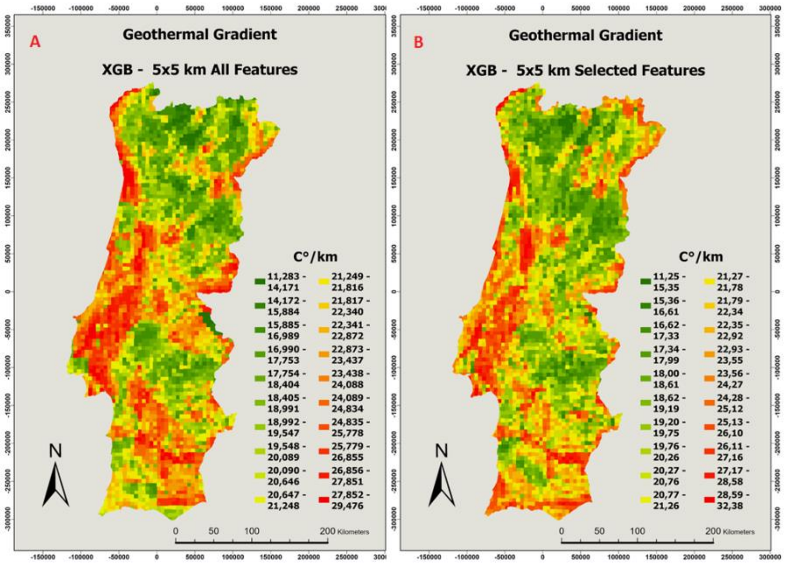
- **Values range: 32,78 – 104,96 mW/m²**



Results

- ## XGB Models:

Geothermal Gradient				
	XGB (5x5 grid)		XGB (2.5x2.5 grid)	
Hyperparameters	All Features	Selected from Model	All Features	Selected from Model
Random State	2	2	2	2
Learning Rate	0,1	0,05	0,01	0,05
N Estimators	500	100	500	100
Max Depth	15	12	5	5
Gamma	10	10	10	10
Reg Lambda	10	1	0,1	0,1
Errors				
MAE	3,787958992	3,614766839	4,356813204	4,275029737
RMSE	4,812866168	4,778081377	5,420207974	5,340590027
R2	0,14861505	0,204343513	-0,069126467	-0,043858095

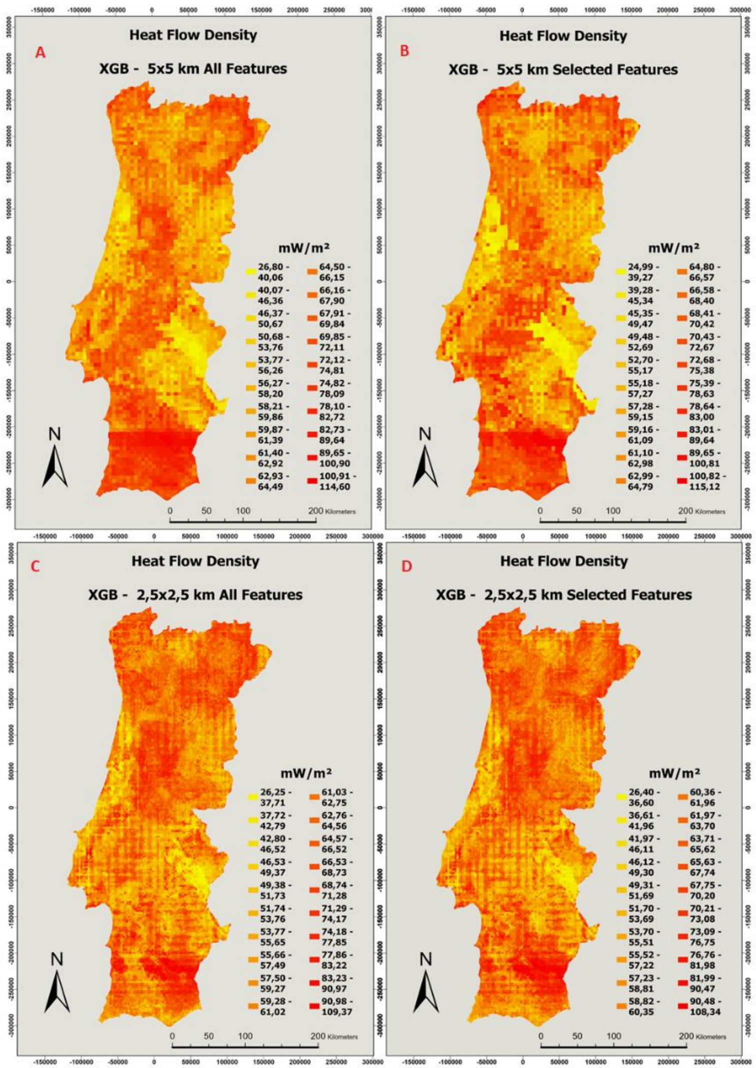


- **Values range: 11,25 – 32,38 °C/km**

Results

- ## XGB Models:

Heat Flow Density				
	XGB (5x5 grid)		XGB (2.5x2.5 grid)	
Hyperparameters	All Features	Selected from Model	All Features	Selected from Model
Random State	2	2	2	2
Learning Rate	0,1	0,01	0,05	0,01
N Estimators	200	500	100	500
Max Depth	5	5	5	5
Gamma	10	0,1	0,1	10
Reg Lambda	0,1	0,1	1	1
Errors				
MAE	12,26850403	12,02276277	12,91684395	12,93098171
RMSE	15,71138418	15,30638079	16,39341767	16,63282622
R2	0,170151418	0,220288661	0,209721412	0,194997342



- Values range: 24,99 – 115,12 mW/m²

Discussion

- **Random Forests performed slightly better than XGB**
  - **Lower RMSE**
  - **Overall, lower MAE**
  - **Higher R^2**

	Geothermal gradient		Heat flow density	
5x5 grid:	RF	XGB	RF	XGB
MAE	3,669	3,615	12,068	12,023
RMSE	4,707	4,778	15,033	15,306
R2	0,215	0,204	0,248	0,220
2,5x2,5 grid:	RF	XGB	RF	XGB
MAE	4,017	4,275	12,928	12,931
RMSE	5,138	5,341	16,095	16,633
R2	0,045	-0,044	0,232	0,195

- **Better results achieved with data aggregated into the lower spatial resolution grid (5x5 km)**
  - **This could be due to the presence of data sources with low spatial resolution used as features in the modelling**

Discussion

- The values of RMSE are relatively high and indicate that models' generalization capacity is not too strong

- However, the models were able to predict the overall spatial distribution of low and high values correctly

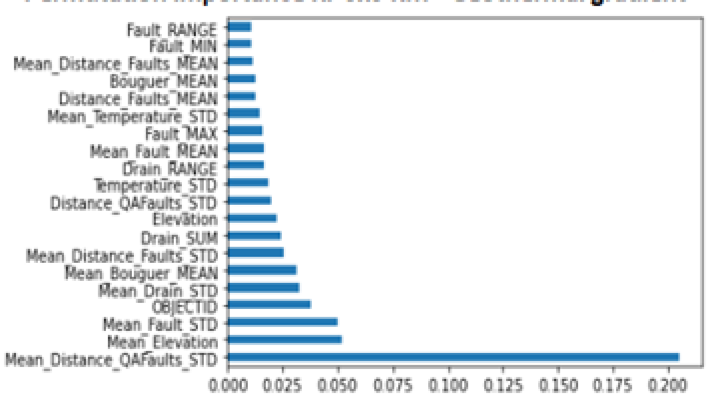
- MAE suggests the maps were able to detect potential zones. Lowest MAE predicting geothermal gradient and surface heat flow density was 3,61 C°/km and 12,02 mW/m^2, respectively

- As an example, Shahdi et al., (2021) achieved:
  - **MAE scores of 5,6 and 7 (C°/km) with XGB and DNN, respectively. (Comparing predicted geothermal gradient and temperature at specific depths)**
  - **MAE score of 3,21 °C and RMSE of 4,94 with XGB, and MAE of 3,25 °C and RMSE of 5.01 with RF. (When evaluating the models with their main dataset to predict subsurface temperatures)**

Discussion

- ## Geothermal Gradient:
  - **Neighbors Std of distance to quaternary faults was given the highest score in both importance calculations (feature importance and permutation importance)**
  - **Among the most important features, the criteria were:**
    - **distance to quaternary faults**
    - **elevation**
    - **density and distance to faults**
    - **temperature**
    - **Bouguer anomaly**
    - **density of drainages**

Permutation importance RF 5x5 KM - Geothermal gradient

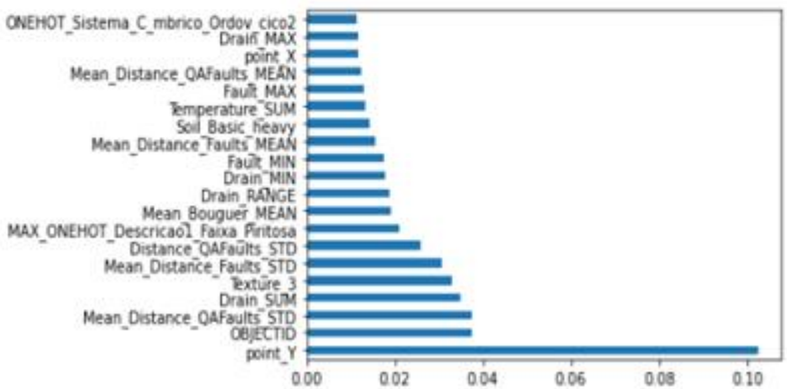


Permutation Importance	
Random Forests 5x5 km Geothermal Gradient	
Weight	**Feature**
0.2288 ± 0.0849	Mean_Distance_QAFaults_STD
0.0610 ± 0.0349	Mean_Elevation
0.0539 ± 0.0074	Mean_Fault_STD
0.0428 ± 0.0154	OBJECTID
0.0338 ± 0.0111	Mean_Bouguer_MEAN
0.0319 ± 0.0048	Mean_Drain_STD
0.0281 ± 0.0118	Mean_Distance_Faults_STD
0.0254 ± 0.0072	Drain_SUM
0.0229 ± 0.0038	Drain_RANGE
0.0208 ± 0.0093	Elevation
0.0202 ± 0.0064	Distance_QAFaults_STD
0.0186 ± 0.0093	Temperature_STD
0.0168 ± 0.0065	Fault_MAX
0.0148 ± 0.0044	Mean_Temperature_STD
0.0137 ± 0.0086	Mean_Fault_MEAN
0.0136 ± 0.0053	Bouguer_MEAN
0.0135 ± 0.0065	Mean_Distance_Faults_MEAN
0.0131 ± 0.0026	Fault_RANGE
0.0122 ± 0.0045	Distance_Faults_MEAN
0.0122 ± 0.0036	Fault_STD

Discussion

- **Surface heat flow density:**
  - **Latitude had highest importance score in both importance calculations**
  - **Neighbors Std of distance to quaternary faults also had great importance**
  - **The most important features were:**
    - latitude
    - distance to quaternary faults
    - distance and density of faults
    - density of drainages
    - soil texture (coarse)
    - soil composition (basic heavy)
    - lithology (Pyrite belt)
    - geochronology systems (Cambrian Ordovician)
    - Bouguer anomaly
    - temperature

Permutation importance RF 5x5 KM - Heat flow density



Permutation Importance	
Random Forests 5x5 km Heat Flow Density	
Weight	**Feature**
0.0977 ± 0.0205	point_Y
0.0429 ± 0.0107	Mean_Distance_QAFaults_STD
0.0393 ± 0.0130	OBJECTID
0.0320 ± 0.0045	Drain_SUM
0.0302 ± 0.0083	Texture_3
0.0292 ± 0.0085	Distance_QAFaults_STD
0.0270 ± 0.0044	Mean_Distance_Faults_STD
0.0215 ± 0.0039	Drain_RANGE
0.0209 ± 0.0065	MAX_ONEHOT_Descricao1_Faixa_Piritosa
0.0178 ± 0.0069	Fault_MIN
0.0168 ± 0.0027	Drain_MIN
0.0167 ± 0.0060	Mean_Bouguer_MEAN
0.0133 ± 0.0048	Mean_Distance_Faults_MEAN
0.0119 ± 0.0020	Fault_MAX
0.0118 ± 0.0026	Mean_Distance_QAFaults_MEAN
0.0118 ± 0.0034	Mean_Drain_STD
0.0116 ± 0.0037	Temperature_SUM
0.0115 ± 0.0036	Soil_Basic_heavy
0.0112 ± 0.0036	Drain_MAX
0.0110 ± 0.0045	ONEHOT_Sistema_C_mbrico_

Discussion

- Geological, hydrogeological and geophysical parameters were the most important features
- Surface heat flow density is more influenced by surface properties than geothermal gradient. Soil texture and composition, and lithology (distinct thermal conductivities) are related parameters
- The presence of high values of heat flow density in the south of Portugal and in the Pyrite-belt could be one of the factors that contributed to the great importance given to latitude

Discussion

- Geothermal gradient is more related to subsurface parameters

- However, they could be linked to the presence of faults, especially recent (quaternary) faults, high density of drainages and lower Earth's crust thickness (Bouguer anomaly)

- In the prediction maps, the influence given to quaternary faults can be seen by higher values close to where there are fault traces

Discussion

- **Limitations and future work:**

    - Only CV, No extra dataset for validation
    - Errors associated with estimations and measurements
    - More errors where there are no training data

    - Improving the accuracy of the models is likely to be achieved with:
        - More and better spatially distributed training data
        - Fitting into the model other important features related to geothermal parameters:
            - Physical and chemical parameters of the ground
            - Remote sensing images such as thermal infra red, radar and multispectral images
        - Test other ML models (DNN, CNN)
        - Comparisons with geo statistics
            - However, it is not able to be used for modeling

Conclusion

- Two ensemble machine learning algorithms to predict crucial thermal parameters of the surface and subsurface of Mainland Portugal
- Prediction maps were able to target regions with higher geothermal potential and can be applied to aid geothermal exploration surveys
- Important features were identified and confirmed the relevance of geological, hydrogeological and geophysical data, especially distance to quaternary faults
- Comparing RF and XGB revealed slightly better results overall with RF
- Comparing the grids revealed better achievements with lower spatial resolution (5x5 km)

NOVA IMS Information Management School

Bibliographic References

- Ammann, M. W., Walker, A. M., Stackhouse, S., Wookey, J., Forte, A. M., Brodholt, J. P., & Dobson, D. P. (2014). Variation of thermal conductivity and heat flux at the Earth's core mantle boundary. Earth and Planetary Science Letters, 390, 175–185. https://doi.org/10.1016/j.epsl.2014.01.009
- Arola, T., Korhonen, K., Martinkauppi, A., Leppäharju, N., Hakala, P., Ahonen, L., & Pashkovskii, M. (2019). Creating shallow geothermal potential maps for Finland using finite element simulations and machine learning. European Geothermal Congress 2019, June, 6.
- Assouline, D., Mohajeri, N., Gudmundsson, A., & Scartezzini, J. L. (2019). A machine learning approach for mapping the very shallow theoretical geothermal potential. Geothermal Energy, 7(1). https://doi.org/10.1186/s40517-019-0135-6
- Badino, G. (2005). Underground drainage systems and geothermal flux. Acta Carsologica, 34(2), 277–316. https://doi.org/10.3986/ac.v34i2.261
- Barbier, E. (2002). Geothermal energy technology and current status: An overview. Renewable and Sustainable Energy Reviews, 6(1–2), 3–65. https://doi.org/10.1016/S1364-0321(02)00002-3
- Bignall, G., Milicich, S., Ramirez, E., Rosenberg, M., Kilgour, G., & Rae, A. (2010). Geology of the Wairakei-Tauhara Geothermal System, New Zealand. Proceedings World Geothermal Congress, November 2019, 25–29.
- Booysen, R., Gloaguen, R., Lorenz, S., Zimmermann, R., & Nex, P. A. M. (2021). Geological Remote Sensing. In Encyclopedia of Geology (2nd ed.). Elsevier Inc. https://doi.org/10.1016/b978-0-12-409548-9.12127-x
- L. Breiman, "Random forests", Mach. Learn., vol. 45, no.1, pp. 5-32, Oct. 2001. doi:10.1023/A:1010933404324
- Chamorro, C. R., García-Cuesta, J. L., Mondéjar, M. E., & Linares, M. M. (2014). An estimation of the enhanced geothermal systems potential for the Iberian Peninsula. Renewable Energy, 66, 1–14. https://doi.org/10.1016/j.renene.2013.11.065
- Corniello A., D. D. (2012). Hydrogeology key role in a geothermal exploration in southern Italy. January. http://www.researchgate.net/publication/255486060_Hydrogeology_key_role_in_a_geothermal_exploration_in_southern_Italy
- Coro, G., & Trumpy, E. (2020). Predicting geographical suitability of geothermal power plants. Journal of Cleaner Production, 267, 121874. https://doi.org/10.1016/j.jclepro.2020.121874
- Fahil, A. S., Ghoneim, E., Noweir, M. A., & Masoud, A. (2020). Integration of well logging and remote sensing data for detecting potential geothermal sites along the gulf of Suez, Egypt. Resources, 9(9). https://doi.org/10.3390/RESOURCES9090109
- Gangwani, P., Soni, J., Upadhyay, H., & Joshi, S. (2020). A Deep Learning Approach for Modeling of Geothermal Energy Prediction. International Journal of Computer Science and Information Security, 18(1), 62–65.
- González, D. L., & Rodríguez-Gonzálvez, P. (2019). Detection of geothermal potential zones using remote sensing techniques. Remote Sensing, 11(20). https://doi.org/10.3390/rs11202403
- Han, M., Feng, Y., Zhao, X., Sun, C., Hong, F., & Liu, C. (2019). A Convolutional Neural Network Using Surface Data to Predict Subsurface Temperatures in the Pacific Ocean. IEEE Access, 7, 172816–172829. https://doi.org/10.1109/ACCESS.2019.2955957
- Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. Artificial Intelligence Review, 26(3), 159–190. https://doi.org/10.1007/s10462-007-9052-3
- Lai, J. P., Chang, Y. M., Chen, C. H., & Pai, P. F. (2020). A survey of machine learning models in renewable energy predictions. Applied Sciences (Switzerland), 10(17). https://doi.org/10.3390/app10175975
- Lerche, I. (1991). Temperature dependence of thermal conductivity and its impact on assessments of heat flux. Pure and Applied Geophysics PAGEOPH, 136(2–3), 191–200. https://doi.org/10.1007/BF00876371
- Lund, J. W., & Freeston, D. H. (2001). World-wide direct uses of geothermal energy 2000. Geothermics, 30(1), 29–68. https://doi.org/10.1016/S0375-6505(00)00044-4
- Merembayev, T., Yunussov, R., & Yedilkhan, A. (2019). Machine learning algorithms for classification geology data from well logging. 14th International Conference on Electronics Computer and Computation, ICECCO 2018, 9–13. https://doi.org/10.1109/ICECCO.2018.8634775
- Mock, J. E., Tester, J. W., & Wright, P. M. (1997). Geothermal energy from the earth: Its potential impact as an environmentally sustainable resource. Annual Review of Energy and the Environment, 22(1), 305–356. https://doi.org/10.1146/annurev.energy.22.1.305
- Mohamed, H. S., Abdel Zaher, M., Senosy, M. M., Saibi, H., el Nouby, M., & Fairhead, J. D. (2015). Correlation of Aerogravity and BHT Data to Develop a Geothermal Gradient Map of the Northern Western Desert of Egypt using an Artificial Neural

Bibliographic References

- Network. Pure and Applied Geophysics, 172(6), 1585–1597. https://doi.org/10.1007/s00024-014-0998-1
- Nguyen, A., Pasquier, P., & Marcotte, D. (2017). Borehole thermal energy storage systems under the influence of groundwater flow and time-varying surface temperature. Geothermics, 66, 110–118. https://doi.org/10.1016/j.geothermics.2016.11.002
- Olasolo, P., Juárez, M. C., Morales, M. P., Damico, S., & Liarte, I. A. (2016). Enhanced geothermal systems (EGS): A review. Renewable and Sustainable Energy Reviews, 56, 133–144. https://doi.org/10.1016/j.rser.2015.11.031
- Omwenga, B. M. (2020). Geothermal Well Site Suitability Selection Using Geographic Information Systems (GIS) and Remote Sensing: Case Study of the Eburru Geothermal Field. 45th Workshop on Geothermal Reservoir Engineering, 1, 1–6.
- F. Pedregosa et al., Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825–2830, 2011.
- Pukelsheim. (1994). The three sigma rule. American Statistician, 48(2), 88–91. https://doi.org/10.1080/00031305.1994.10476030
- Ramalho, E. C. (2014). O papel do atlas geotérmico nacional no fomento da exploração da energia geotérmica em Portugal continental. Comunicacoes Geologicas, 101, 833–836.
- Ramalho, E., Correia, A., 2015. Atlas Geotérmico de Portugal Continental [mapa] https://geoportal.lneg.pt/mapa/?mapa=AtlasGeotermico [20 Fev, 2022].
- Rezvanbehbahani, S., Stearns, L. A., Kadivar, A., Walker, J. D., & van der Veen, C. J. (2017). Predicting the Geothermal Heat Flux in Greenland: A Machine Learning Approach. Geophysical Research Letters, 44(24), 12,271-12,279. https://doi.org/10.1002/2017GL075661
- Robertson, E. C. (1988). Thermal Properties of Rocks. US Department of the Interior: Geological Survey, 88–441.
- S. Angra and S. Ahuja, “Machine learning and its applications: A review,” in 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC). IEEE, 2017, pp. 57–60.
- Shahab, A., & Singh, M. P. (2019). Comparative analysis of different machine learning algorithms in classification of suitability of renewable energy resource. Proceedings of the 2019 IEEE International Conference on Communication and Signal Processing, ICCSP 2019, 2013, 360–364. https://doi.org/10.1109/ICCSP.2019.8697969
- Shahdi, A., Lee, S., Karpatne, A., & Nojabaei, B. (2021). Exploratory analysis of machine learning methods in predicting subsurface temperature and geothermal gradient of Northeastern United States. Geothermal Energy, 9(1). https://doi.org/10.1186/s40517-021-00200-4
- Sun, Z., Jiang, B., Li, X., Li, J., & Xiao, K. (2020). A data-driven approach for lithology identification based on parameter-optimized ensemble learning. Energies, 13(15), 1–15. https://doi.org/10.3390/en13153903
- Suwai, J. (2009). Basic Hydrogeology in Geothermal Systems. Short Course V on Exploration for Geothermal Resources, 1–9.
- Tester, J. (2007). Congressional Testimony for. 1–13.
- Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 785–794.
- Tinti, F., Kasmaee, S., Elkarmoty, M., Bonduà, S., & Bortolotti, V. (2018). Suitability evaluation of specific shallow geothermal technologies using a GIS-Based multi criteria decision analysis implementing the analytic hierarchic process. Energies, 11(2). https://doi.org/10.3390/en11020457
- Trumpy, E., Donato, A., Gianelli, G., Gola, G., Minissale, A., Montanari, D., Santilano, A., & Manzella, A. (2015). Data integration and favourability maps for exploring geothermal systems in Sicily, southern Italy. Geothermics, 56(January 2021), 1–16. https://doi.org/10.1016/j.geothermics.2015.03.004
- van der Meer, F., Hecker, C., van Ruitenbeek, F., van der Werff, H., de Wijkerslooth, C., & Wechsler, C. (2014). Geologic remote sensing for geothermal exploration: A review. International Journal of Applied Earth Observation and Geoinformation, 33(1), 255–269. https://doi.org/10.1016/j.jag.2014.05.007
- Zhang, L., & Zhan, C. (2017). Machine Learning in Rock Facies Classification: An Application of XGBoost. January, 1371–1374. https://doi.org/10.1190/igc2017-351

NOVA IMS Information Management School

Methodology

- RF and XGB Regressor were utilized to predict both labels
- Scikit Learn Python Module and Xgboost library

- Modelling and evaluation:
  - 5-fold Cross Validation
  - R^2, RMSE, and MAE
  - Grid Search CV – find best hyperparameters
  - Feature Selection with Features importance and Select from Model
  - Permutation Importance

- Show Jupyter Notebook

ML with Geospatial Data

Suggestions:

- GIS (ArcGIS, QGIS) – can be used to preprocess data and for visualization
- Python – preprocess and ML modelling
  - Pandas and Numpy
  - Matplotlib, Seaborn, Plotly
  - Scikit-Learn
    - Feature engineering
    - Feature selection
    - Dimensionality reduction
    - ML models
    - ETC
  - Keras and Tensor flow - deep learning
  - Geospatial libraries
- Other programming languages and libraries
- Softwares such as weka, orange, etc

Thank you!

Acreditações e Certificações da NOVA IMS

UNIGIS Erasmus Mundus Programme Education and Culture DG A3ES Double Degree Master Course in Information Systems Management iSchools eduniversal EMOS ABET Computing Accreditation Commission USGIF

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa