Proceedings World Geothermal Congress 2020+1
Reykjavik, Iceland, April - October 2021

Automation on Enhancing the Ground Temperature Sensing Using Machine Learning Approach

Avish Mehta[1] Kanish Shah[2] Manan Shah[3]

[1,3] Department of Chemical Engineering, School of Technology, Pandit Deendayal Petroleum University, Gandhinagar-382007, Gujarat, India

[2]Department of Computer Engineering, Indus University, Ahmedabad, Gujarat, India

Email id: avish.mch19@sot.pdpu.ac.in keshah24@gmail.com Manan.shah.@spt.pdpu.ac.in

Keywords: - Machine Learning, Support Vector Machine, Decision Trees, Geothermal energy, Ground Temperature

ABSTRACT

Geothermal Energy is the inexhaustible and clean form of energy derived from the earth's crust for different purposes like generation of electricity. Water or steam in then used to bring the energy to the earth's surface. The ground temperature is predicted on various factors such as weather and terrain variables such as latitude, longitude ground aspect and slope, monthly precipitation, sunshine duration, snow depth and air temperature. The challenges faced are due to seasonal weather variations and at the depth of 1-2 m significant changes occur in the ground temperatures. After the collection and pre-processing of the data, it can be used to train and deploy the Machine learning model. By using various algorithms such as Decision trees and Support vector Machine we can predict the ground temperature at the unknown locations. By using the help of various accuracy functions such as root mean square method (RMSE) or normalized root mean square error method (NRMSE) we perform the estimations and calculate the efficiency of the model. The model can also be used to show the ground temperature maps and can also the variations of the ground temperature as compared to the previous years and can predict the future ground temperatures. If the data pre-processing and the prediction is done accurately it can be implemented on a larger scale and thus increase the geothermal energy productions.

1. INTRODUCTION

The energy demand by the world population is increasing day by day due to expansion of economy and the population growth with advancements in the energy-intensive technologies (Patel et al., 2020). Fossil fuels are not present in the unlimited quantity and will get over after few years and with the increasing concerns such as global warming, CO_2 level in the atmosphere and the carbon footprint, we need to move towards an alternative source of clean and sustainable form of energy. Geothermal energy is the most versatile form of renewable energy source that can be utilized for many purposes, the major use being in the production of energy. Geothermal energy is the renewable form of energy that is derived from the earth's surface. It is found in the deep surfaces of the earth and it major involves the use of high temperatures produced beneath the earth to generate electricity from the heated water. Use of geothermal energy as an energy source has many advantages such as wide resource coverage, less pollution caused to the environment and thus reduces the carbon footprint, the amount of CO_2 released into in the atmosphere is approximately 50% less than the amount of CO_2 released into the atmosphere by burning of fossil fuels and high reliability (Kauffman & Lee, 2013).

Recent developments and the advancements in the field of machine learning and artificial intelligence using artificial neural networks and convolutional neural networks, has helped in the better accurate estimation of the energy, values and environmental modelling (Jani et al., 2020; Patel et al., 2020; Gupta et al., 2020). Based on the given input dataset a model is build and trained with the help of various algorithms such as decision trees, random forest etc (Shah et al., 2020). and then we do the prediction of various parameters. (Assouline et al., 2019; Patel et al., 2020; Shah et al., 2020) It has been widely used in the lowering of the risk factors associated with the drilling techniques such as prediction of the seismic activities, it has been widely used in the stages of the exploration and development purposes. From the past years machine learning has been used for the geospatial modelling of multiple environmental variables, including solar radiation and wind speed, forecasting of solar radiation over horizontal and tilted surfaces , and short-term forecasting of wind speed and wind power prediction. Machine learning can be used to significantly reduce the cost of the operations and can help in automating the processes, which thus helps to increase the efficiency of the system (Parekh et al., 2020). It can also help in optimizing the extraction of the energy from the earth in terms of efficiency and the safety. Machine learning has also been used in building contour maps such as thermal conductivity contour map of any area based on the data given. Machine learning can help you take decisions by analysing all the factors and can help you take a better decision.

This paper mainly highlights the prediction of the ground temperature-using machine learning which can be further applied to harness the geothermal energy for various different purposes such as electricity generation etc. we integrate machine learning with geothermal for the prediction of the ground temperature. We try to give various algorithms, which can be used for the prediction of the ground temperature using machine learning.

2. LIMITATIONS OF CONVENTIONAL GEOTHERMAL METHODS FOR GROUND TEMPERATURE PREDICTION

Various methods and calculations were done in the previous years for the calculation of the ground temperature.

Takao Katsura et.al (2007) proposed the method for Calculation of ground temperature for heat extraction via multiple ground heat exchangers. In this method, the authors used the superposition principle to the temperature response of cylindrical heat source as well as for a line heat source for the calculation of ground temperatures. For the calculation purposes various assumption were made such as soil is treated to be a isotropic constant solid and the vertical ground heat exchanger (borehole or pile) is treated as a hollow cylinder in the infinite solid (Katsura et al., 2008), gradient of the temperature response Is constant .the main condition for the using of superposition principle is that the temperature response for the cylindrical heat source can be regarded as equivalent to the temperature response for the line heat source when the non-dimensional distance is larger than 10 (L.R. Ingersoll et al., 1948). Various results were determined from the mathematical equations. The absolute error of the average temperature during the extraction period is less than 0.01. Difference between temperatures for the single ground heat exchanger and the multiple ground heat exchangers becomes larger as time elapses. When the ground temperature goes below 0 C, the method does not consider the latent heat generated by soil freezing. The method can also demonstrate the difference of temperature of the ground surrounding each heat exchanger and temperature distribution. After a certain period, the temperature of the ground surrounding the inner heat exchangers was lower than the temperature of the ground surrounding the outer heat exchangers

(Gwadera et al., 2017) proposed the mathematical equations for the calculation of the ground temperature keeping considering the various factors such as the solar energy absorbed by the ground, long-wave radiation heat flux and the evaporation of the water contained in the ground. For the calculation of the average convective fluxes they assumed that the environment is constant with respect to time (Krarti et al., 1995; Mihalakakou et al., 1997). While taking into consideration the solar radiation and the long- wave radiation, the value depends on the various factors such as the degree of the radiation absorption by the ground which further depends on the grounds humidity and the cover, (Duffie & Beckman, 2013) the temperature of the sky is variable during the day and the night and for approximate calculations that the average long-wave radiation flux amounts to 63 W/m2. (Gwadera et al., 2017) While taking into the consideration of evaporation of water from the surface of the ground, certain assumption were taken such as for calculations of the yearly average flux, an averaged value of the saturated vapour pressure was assumed, at the temperature of the surface of the ground and at the ambient temperature. (McAdams, 1954) Various result were drawn from the equation such as The convective heat transfer coefficients between the surface of the ground and the environment are difficult to determine precisely as various empirical dependencies are used for their determination, ambient temperature the temperature of the surface of the ground and the solar radiation flux absorbed by the ground change periodically in the yearly cycle etc. from the above we can interpret that in practical situations we cant assume the above factors constant as their changes will affect the temperature of the ground.

(Kalogirou et al., 2012) performed the generation of the geothermal contours of temperature at various depths including 20m, 50m , and 100m n Cyrus. The complete data was divided into training and validation on a random basis. The author used the neural networks to predict the temperature at a random point by drawing the contour maps of temperature. The prediction error was confirmed to be less than 1.74 degree Celsius. The neural networks were used for the prediction accuracy on 112 pieces of the data and only 4 points showed 5% more deviation than their actual values. The method proposed involves lithology class, elevation, ambient temperatures and rainfall in drawing of the geothermal maps and it realistically produced valid maps of geothermal temperature at the various depths.

From the various methods described in the above papers, there are certain issues with them and for every mathematical formulation various assumptions are made that vary differently in the practical world. Due to seasonal variations and changes in the environmental conditions of the ground, the surroundings, and in the composition of the ground, determination of the ground temperatures with the help of mathematical equations are difficult as conditions may vary from the time to time and from place to place. To overcome this situation we need the help of technology to predict the ground temperatures precisely and accurately.

3. MACHINE LEARNING ALGORITHMS

For the prediction of ground temperature, the use of Decision Tree Regression and Support Vector Regression algorithms are being employed. Due to their better handling of data and less cost along with their robust nature, they have been selected for handling these high dimensional data in order to get better output.

3.1 Decision Trees Regressors

This algorithm is quite efficient in traversing the datasets and completely follows a tree-like path. The root node is the main part in structure of decision tree which decides the splitting in a decision tree. The working of this algorithm goes in the following manner

Algorithm: To_Generate_Decision_Tree, To generate a decision tree from training tuples (denoted by T) of data

Input:

- Data Splitting, Ds, which is a combination of T along with their associated class labels.
- feature_list, denotes the set of features in data
- feature_selection_method, this method defines the splitting criteria which can possibly define the splitting point

Output: Decision tree

Steps in generation of Decision tree

1. create a node N_j
2. if tuples in Ds belong to the same class

return N_j labelled with that class

3. if feature_list is empty then:

return N_j as a leaf node labelled with the majority class in Ds

4. apply feature_selection_method to find the best criteria for split

5. label N_j with the splitting criteria

6. if splitting is multiway and the attribute is distinct

feature_list feature_list – attribute

7. for each output i of splitting criteria

in Ds let Ds_i be the set of tuples satisfying the output i

if Ds_i is empty then

attach a labelled leaf to N with the majority class Ds

else attach the node returned by To_Generate_Decision_Tree to N_j

end for

8. return N_j

The feature selection method plays a key role in deciding the selecting criteria and how the data will be partitioned such that the data goes in to their respective classes. The feature that has the best score for measure is chosen as the splitting feature. So, once the decision tree is built, there is a probability of anomalies that gets reflected in the branches due to outliers in the training data. That's where the concept of pruning comes in to play. This method uses statistical measures and removes the less reliable branches from the decision tree. This method is less complex and much faster at correctly classifying the independent data.

3.2 Support Vector Regressor

This algorithm mainly focuses on finding an appropriate line or a hyperplane (for higher dimensions) by finding out how much error is acceptable in our model. It follows a similar principle as the Support Vector Machine (SVM) with a very minor difference that in the case of regression, a margin of tolerance named epsilon is set here. This margin of tolerance factor's main goal is to individualize the hyperplane which maximises the margin by tolerating the error. Tuning the epsilon helps achieve the desired accuracy of the model.

So, Support Vector Regression works completely on the decision boundaries and hyperplanes. The main aim of this regression algorithm is to map the real numbers coming from training sample to an input domain.

As shown in Fig. 1, there are points marked which depicts the training samples. If these points are inside the decision boundary then it is fair enough to say that the model is fitting the data well. So, the main aim of SVR is to find out the best fit line (hyperplane) where maximum points lie near to it or on it.

So, let us understand the mathematics behind the decision boundaries. Consider these 2 decision boundary lines at a distance d from hyperplane. Consider them at a distance of +d and –d from hyperplane.

Let us assume the equation of hyperplane as:-

$Y = mx + c$

So, the equation of decision boundaries can be :-

$mx + c = + d$

$mx + c = - d$

Hence, any hyperplane that is being constructed should satisfy the SVR as:

$- d < Y - mx + c < +d$

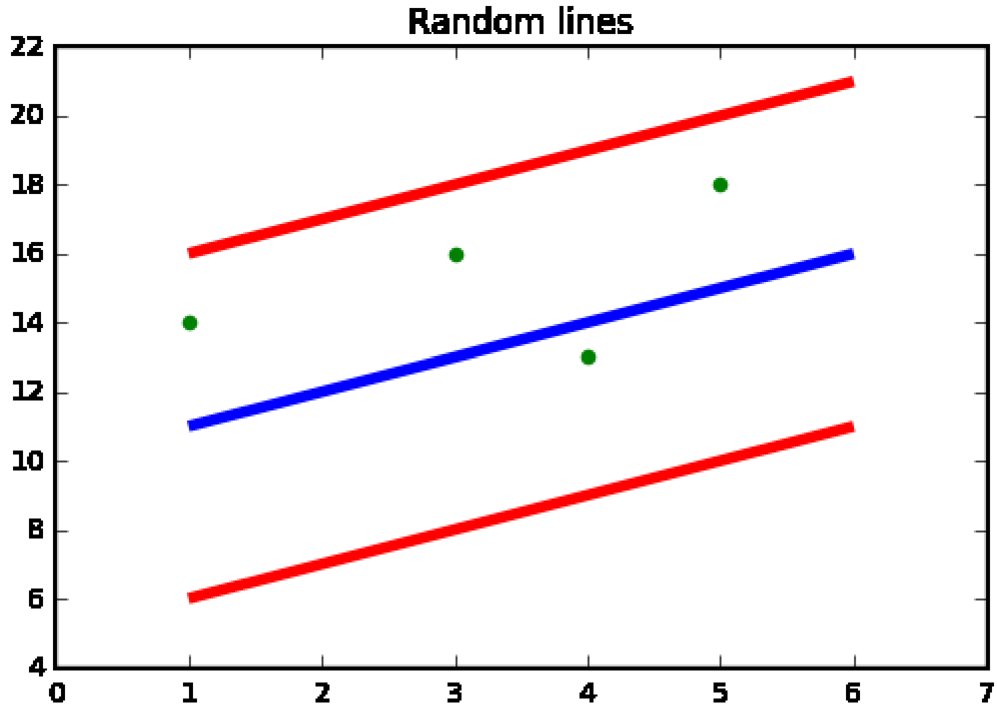

Fig. 1 Diagram showing decision boundaries (marked in Red color) and Hyperplane (marked in Blue color).

4. METHODOLOGY

4.1 Data Pre-processing and Feature Selection

The authors have decided to choose the data for Indian region based on different atmospheric factors such as sunshine duration, precipitation, temperature, cloud cover, ground temperature, air temperature, ground slope and soil moisture. So, our Support Vector Machine and Decision tree models will be trained using the pixel ground temperature values as training labels. The weather variables mentioned above will be considered as the training features. The seasonal variations play a major role in affecting the ground temperature. We have decided to consider shallow horizontal ground loop collectors which are at a depth of 1-2 m because of the lack of proper sediment cover at a particular location. The estimation is being done on an hourly ground temperature time series data and on the basis of that we plan to predict the monthly ground temperature. So, these features are being fed in to these two models separately and try to predict the ground temperature.

4.2 Implementation of Regression models

Once the data gets gathered and pre-processed in a proper way we will implement the Support Vector Regression and Decision tree regression models on the dataset. In the beginning we decide to split the dataset in to training and testing where training consists of 75% of the data and testing consists of 25% of the data. Then we plan to use pipeline in our machine learning model. So, it works by enabling the sequence of data that needs to be transformed and correlate in the model that's needs to be evaluated for achieving the outcome (Shah et al., 2020). Simultaneously the pipelining will be applied to both the models because implementing pipeline really improves the efficiency of the model. Moreover, pipelining also helps in handling the overfitting problem and helps in tuning the hyper-parameters too. After pipelining it is time to implement the regression models. The models get trained based on the input labels provided and at the end the ground temperature gets predicted. The ground temperature data that comes out of prediction then can be compared with the actual ground temperature values in order to find out the difference between the actual and predicted values. Once the model is being implemented, the program will be stored in the disk with help of pickle so in the future if one needs to use it then they can work it with ease.

4.3 Results

Once the models gets trained then we will try to show and display different parameters for the ground temperature prediction such as:-

1. Accuracy:- This parameter gives us the ratio of observations that are corrected predicted to the total number of observations.

2. Precision:- Precision tell us the ratio of observations that are predicted positively correct to the total number of observations predicted positively.

3. F1-score:- It is the harmonic average of recall and precision

4. Testing errors such as Root Mean Square Error (RMSE) and Root Mean Square Deviation (NRMSE):- RMSE values are a form to measure accuracy as it tries to compare the prediction errors of a particular variable with the actual values of that variable.

5. Ground temperature map for each month and at each depth

So, the above mentioned parameters will be then displayed based on how well the model has trained and also how better the model has fit the data. Higher the values of accuracy, precision, f1-score and lesser the mean square errors gives a much better prediction of the ground temperature.

5. CHALLENGES AND FUTURE SCOPE

The main challenge for this proposal is the collection of data for feeding it in to the machine learning model (Naik et al., 2020). As mentioned previously that we have considered various parameters such as sunshine duration, precipitation, temperature, cloud cover, ground temperature, air temperature, ground slope and soil moisture in our dataset. So, there are chances that the collection of data might not be uniform at all the places. As we plan to collect the data from different parts of India so the collection of data can be a challenge. As there are different atmospheric conditions in different parts of India like for instance the sunshine is more in some parts of India like Gujarat, Rajasthan etc. and low for higher altitudes regions like Jammu & Kashmir, eastern parts of India etc. Temperature also varies from states to states of India and similarly all the parameters changes from one place to another. Ground temperature, soil conditions also varies from states to states like for instance Rajasthan has different soil conditions compared to South India states and regions. So, in a nutshell the data gathering is a main challenge that one can face while implementing this methodology. Another important challenge is the synchronization of data with that of machine learning algorithms. There are many instances that the data needs to be cleaned in such a manner so that it becomes feasible for the model to fit it. There might be instances where one cannot get proper data from some regions. So, the pre-processing needs to be done in such a way that the important data does not get lost and at the same time accuracy is also not affected from pre-processing. However, this was just a proposal that we planned to do in the future scope. So, there might be chances that we need to do some minor changes when we actually implement our proposal on the real time datasets. But the methodology proposed will be very useful if some other researcher wants to perform the experiment with some modifications. Apart from these algorithms one can even try some other machine learning algorithms such as Random Forest, Lasso regression, deep learning algorithms such as Artificial Neural Network, Convolutional Neural networks etc.

6. CONCLUSION

The current paper is proposing a methodology to predict the ground temperature prediction using the data from different seasonal variations and from different regions of India. The data consists of various atmospheric as well as ground parameters too for the prediction. This paper proposes Support Vector Regression and Decision Tree regression models which describes every aspect of the model in detail by providing the evaluation metrics. The algorithms formulated for the ground temperature prediction are efficient enough to predict the ground temperature in a much efficient way and the accuracy can rise to more than 95% if the steps of the algorithms are properly implemented. The most important evaluation metrics to look out for here is the accuracy and the mean squared errors. Thus, this paper will be useful for the viewers and can act as the guide of incorporation with various other machine learning algorithms as well as to perform a hybrid of geothermal methods as well as machine learning which might enhance the efficiency and effectiveness.

REFERENCES

Duffie, J. A., & Beckman, W. A. (2013). Wiley: Solar Engineering of Thermal Processes, 4th Edition - John A. Duffie, William A. Beckman. http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470873663.html

Gupta, A., Dengre, V., Kheruwala, H.A. Shah, M., (2020). Comprehensive review of text-mining applications in finance. Financ Innov 6, 39 (2020). https://doi.org/10.1186/s40854-020-00205-1.

Gwadera, M., Larwa, B., & Kupiec, K. (2017). Undisturbed ground temperature-Different methods of determination. Sustainability (Switzerland), 9(11). https://doi.org/10.3390/su9112055

Jani K, Chaudhuri M, Patel H, Shah M (2020) Machine learning in films: an approach towards automation in film censoring. J Data, Inf Manag 2(1):55-64. doi:10.1007/s42488-019-00016-9.

Kalogirou, S. A., Florides, G. A., Pouloupatis, P. D., Panayides, I., Joseph-Stylianou, J., & Zomeni, Z. (2012). Artificial neural networks for the generation of geothermal maps of ground temperature at various depths by considering land configuration. *Energy*, *48*(1), 233–240. https://doi.org/10.1016/j.energy.2012.06.045.

Katsura, T., Nagano, K., & Takeda, S. (2008). Method of calculation of the ground temperature for multiple ground heat exchangers. Applied Thermal Engineering, 28(14–15), 1995–2004. https://doi.org/10.1016/j.applthermaleng.2007.12.013

Kauffman, J., & Lee, K. M. (2013). Handbook of Sustainable Engineering. Handbook of Sustainable Engineering, October, 1–1298. https://doi.org/10.1007/978-1-4020-8939-8

Krarti, M., Claridge, D. E., & Kreider, J. F. (1995). Analytical model to predict nonhomogeneous soil temperature variation. Journal of Solar Energy Engineering, Transactions of the ASME, 117(2), 100–107. https://doi.org/10.1115/1.2870823

L.R. Ingersoll et al. (1948). Theory of the ground pipe heat source for heat pump. ASHVE Journal Section.

McAdams, W. H. (1954). Heat Transmission. McGraw-Hill by William H McAdams.

Mihalakakou, G., Santamouris, M., Lewis, J. O., & Asimakopoulos, D. N. (1997). On the application of the energy balance equation to predict ground temperature profiles. Solar Energy, 60(3–4), 181–190. https://doi.org/10.1016/S0038-092X(97)00012-1

Naik, B., Mehta, A., Shah, M., (2020) Denouements of machine learning and multimodal diagnostic classification of Alzheimer's disease. Vis. Comput. Ind. Biomed. Art 3, 26. https://doi.org/10.1186/s42492-020-00062-w.

Parekh, P., Patel, S., Patel, N., Shah, M., Systematic review and meta-analysis of augmented reality in medicine, retail, and games. Vis. Comput. Ind. Biomed. Art 3, 21 (2020). https://doi.org/10.1186/s42492-020-00057-7.

Patel H, Prajapati D, Mahida D, Shah M (2020) Transforming petroleum downstream sector through big data: a holistic review. J Pet Explor Prod Technol 10(6):2601-2611.

Patel, D., Shah, D. & Shah, M. (2020) The Intertwine of Brain and Body: A Quantitative Analysis on How Big Data Influences the System of Sports. Ann. Data. Sci. 7, 1–1. https://doi.org/10.1007/s40745-019-00239-y

Patel, P., Patel, B., Vekaria, E., Shah, M., (2020) Biophysical economics and management of biodiesel, a harbinger of clean and sustainable energy. Int J Energ Water Res 4, 411–423 (2020). https://doi.org/10.1007/s42108-020-00087-0.

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

Shah, K., Patel, H., Sanghvi, D., Shah, M., (2020). A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification. https://doi.org/10.1007/s41133-020-00032-0.

Shah, N., Engineer, S., Bhagat, N., Chauhan, H., Shah, M. (2020). Research Trends on the Usage of Machine Learning and Artificial Intelligence in Advertising. Augmented Human Research, 5(1). doi:10.1007/s41133-020-00038-8.

Shah D, Dixit R, Shah A, Shah P, Shah M (2020) A comprehensive analysis regarding several breakthroughs based on computer intelligence targeting various syndromes. Augment Hum Res 5(1):14. https://doi.org/10.1007/s41133-020-00033-z.