19th Brazilian Congress of Thermal Sciences and Engineering
November 6th–10th, 2022, Bento Gonçalves - RS - Brazil

ENC-2022-0617
ESTIMATING THE HIGHER HEATING VALUE OF CANDIOTA COAL USING DIFFERENT METHODS OF MACHINE LEARNING

Paulo Antônio Ribeiro Morgado
Bruno Roberto Pasa
Nina Paula Gonçalves Salau
Rodolfo Rodrigues
Laboratório de Engenharia de Processos Assistida por Computador - LEPAC
Departamento de Engenharia Química - Universidade Federal de Santa Maria - UFSM
Prédio 9B – Centro de Tecnologia – Av. Roraima, n. 1000 – Camobi – Santa Maria, RS
morgado.paulo@acad.ufsm.br, bruno.pasa@acad.ufsm.br, ninasalau@ufsm.br, rodolfo.rodrigues@ufsm.br

Abstract. *Analyzing coal samples through proximate analysis is a cheaper and less time-consuming method of describing a coal sample composition than the ultimate analysis. It is of interest to try to establish different relationships between coal composition and the higher heating value. Establishing these kinds of relationships, though useful, can sometimes pose some challenges that will require more sophisticated and innovative methods of regression. This work aimed to develop and compare three different methods of machine learning algorithms to develop relationships between the proximate analysis of Brazilian high-ash coal samples and their higher heating value. All three models performed satisfactorily within the limits of acceptable errors, taking into account the heterogeneity of data used in their training. Even in comparison with linear regression, all models performed very similarly, but, only the artificial neural network method was capable of having better metrics than the linear regression.*

Keywords: *Candiota coal, proximate analysis, machine learning, higher heating value*

1. INTRODUCTION

Measuring moisture, ash, volatile matter and fixed carbon contents of coal, i.e., proximate analysis, is a fairly simple and well-known method of analyzing a coal sample (Speight, 2015). This fact in conjunction with the importance of verifying the higher heating value of a coal sample to rule a proper use for it, methods of developing regression models that can establish some relationship between those properties and the higher heating value of coal samples are fairly studied in the area (Akhtar *et al.*, 2017). Some important work exploring nonlinear models for predicting the higher heating value of coal samples should be noted. Akkaya (2009) minutiously explores several nonlinear models establishing relationships between the higher heating value and composition data taken from Turkish coal samples, obtaining a best-fit model with a coefficient of determination (R^2) of 0.97. Akhtar *et al.* (2017) developed two linear based correlation models for predicting the higher heating values of Pakistani coals, which presented values of R^2 of 0.98 and 0.99. An important topic explored by Akhtar *et al.* (2017) is that these kinds of regression models should take into account the geographical region in which that specific sample batch was collected, and thereby, the same regression model can perform differently concerning the coal's composition specificity.

With the recent growth of machine learning-based regression models, some more work regarding the prediction of the higher heating value of coal samples using those types of models emerged. Tan *et al.* (2015) explored a method of using support vector regression to develop correlations between coal composition and higher heating value, for this, two different datasets were used: 167 samples of Chinese coal and 4540 samples of U.S. coal. The first model obtained a correlation coefficient (R) of 0.986, while the second one had an R of 0.996. An important conclusion found in this work is that developing different correlations from coal samples collected in distinct geographical regions results in much more accurate results.

Thus, further studies on the subject of using machine learning regression models to predict the higher heating value of specific coal samples should be conducted. In this work, four machine learning methods were studied to predict the higher heating value of Candiota coal samples based on the ash, volatile matter and fixed carbon contents obtained from the proximate analysis. All of the models showed extremely promising results, being the artificial neural network one

the one that showed the best results when compared to the other three, with an R of 0.919. To the best of the authors' knowledge, no other work on this kind of regression specifically for Candiota coal was developed to date.

2. THEORETICAL BACKGROUND AND METHODOLOGY

2.1 Coal Analysis

2.1.1 Proximate Analysis

Several methods of analyzing coal were developed throughout contemporary history, each with the goal of giving a proper use to a specific type of coal, based on its physicochemical properties. The most common method of analyzing coal is the proximate analysis, which results in the mass percentage of moisture, ash, volatile matter and fixed carbon.

The conventional test procedure for proximate analysis is to calculate the fourth property after determining the first three ones. For the sake of determining the amount of moisture, volatile matter, and ash in a coal sample, heating in a controlled environment for a specific time is used to determine the sample composition. Due to this heating, both moisture loss, and, at higher temperatures, volatile matter loss will induce a weight loss in the sample. After the final temperature has been reached, the combustion residue is ash. Fixed carbon is the difference between the sum of the three known numbers subtracted from 100 (Speight, 2015).

2.1.2 Higher Heating Value

Heating value is a measurement of the heating ability of a specific coal sample. The heating value is needed to estimate the amount of coal needed to produce a desired amount of heat. The calorific value of a coal sample is expressed through the combination of the Gross Calorific Value, called Higher Heating Value (HHV) in this work, and the Net Calorific Value, sometimes called Lower Heating Value (LHV). The latent heat of condensation of the water generated during the sample combustion is the difference between the HHV and the LHV. The HHV indicates that all of the water vapor produced during the combustion process is fully condensed (Speight, 2015). Hereupon, it becomes evident that the HHV is an important metric for coal.

According to Speight (2015), the more traditional method of HHV determination is from the complete burning of the sample in a bomb calorimeter, which is the most suitable and accurate equipment used to determine the calorific value.

2.1.3 Coal Rank

Coal rank is the measure of the degree of organic metamorphism (coalification) of a coal (Stracher *et al.*, 2010). This rank ranges from, ordered from low to high-rank order, peat, lignite, sub-bituminous, bituminous, semi-anthracite and, finally, anthracite coals (O'Keefe *et al.*, 2013). The classification of a coal is an important factor that dictates its application (Ward and Suárez-Ruiz, 2008). For example, low-rank coals are usually used as a fuel in thermal power plants, whereas high-rank coals could be used as a feedstock in industry. In general, with an increase in rank, it's also possible to see a decrease in the percentage of ash and an increase in calorific value.

2.1.4 Characterizing Candiota Coal

The state of Rio Grande do Sul holds around 89% of all the coal currently mined in Brazil, being Candiota coalfield alone responsible for 42% of the state's coal reserves (Kalkreuth *et al.*, 2006). As thoroughly explored by Kalkreuth *et al.* (2013), Candiota coal deposits can be divided into 17 different seams. Of all the 17 seams, two of them, Candiota Superior (CCS) and Candiota Inferior (CCI), can be highlighted as having the most economic importance.

Due to the average high ash content found in Candiota coal samples in Kalkreuth *et al.* (2013) work, those samples were classified, according to the UN-ECE (1998), as very low-quality coals. When analyzing the rank of those same samples, having a mean vitrinite random reflectance of 0.41%, the coals are categorized as sub-bituminous by the same commission. Due to those characteristics, Candiota coal has very limited use and is mainly applied as a fuel in thermal power plants.

The average values, in dry basis, for Candiota coal composition can be found in the old "Companhia Riograndense de Mineração" website under the section called "Jazida de Candiota" (Companhia Riograndense de Mineração - CRM, 2008). The average values are displayed in Table 1.

Table 1. Average values, in dry basis, for Candiota coal composition.

Characteristic	Average Value [%]
Moisture	16.00
Ash	52.70
Volatile Matter	21.20
Fixed Carbon	26.10
Total Sulfur	1.30

2.2 The Dataset

As discussed by Goodfellow *et al.* (2016), the growth of machine learning and deep learning-based models is deeply correlated with data availability and more powerful methods of computing. That being said, when it comes to the engineering field, extremely large datasets are not often available for use. The reason is fairly simple to understand, experimental data requires time and, more important, money to gather. In this work, the dataset used for developing all machine learning models was provided by CRM - Companhia Riograndense de Mineração, a state-owned company responsible for the research, extraction, and trading of Candiota coal. 119 coal samples were analyzed using proximate analysis and the respective ash, volatile matter and fixed carbon contents (in dry basis), as well as the higher heating value, were stored. Unfortunately, the dataset contained coal samples from different seams mixed, which largely contributed to some of the variance found in the dataset itself. Admittedly, the dataset used is relatively small for a machine learning project with a big scope, so, for that reason, all models developed will be limited by the amount of data used. Nevertheless, this sets a precedent for new, more complex and more in-depth studies of Candiota coal using larger datasets. A general view of the dataset, displaying the relationships between coal composition itself, as well as coal composition and HHV, can be seen in Figure 1.

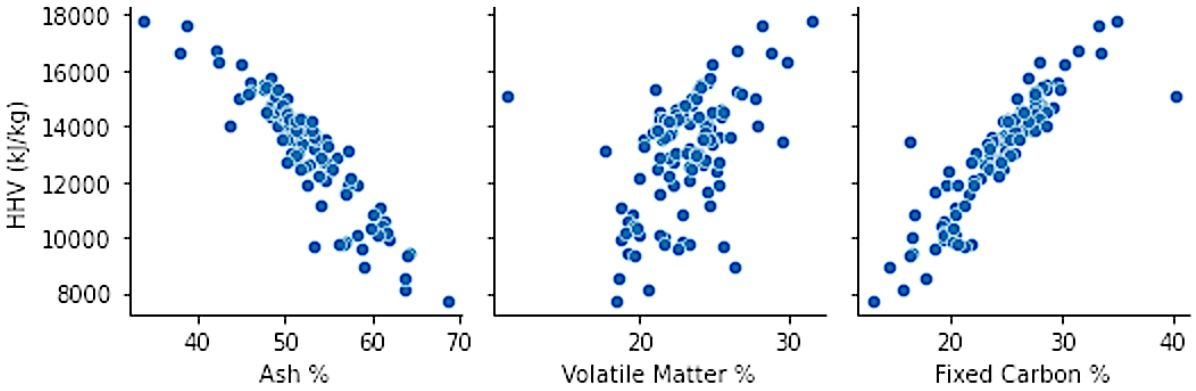

Figure 1. Variables and their correlations.

As it can be seen, the Higher Heating Value shows a high linearity degree with both the Fixed Carbon and Ash. That being said, this relationship does not occur with the Higher Heating Value and Volatile Matter. It is also important to describe the data used to train and test all models studied in this work. As a consequence of the use of machine learning as a regression tool, the models presented in this work are most suited to predict the higher heating value whose coal composition lies within the range used in the training set. Table 2 describes the data range used in this work.

Table 2. Range of values studied in this work.

Values	Ash (%)	Volatile Matter (%)	Fixed Carbon (%)	HHV (kJ/kg)
Minimum	33.75	11.20	13.00	7757.14
Mean	52.50	22.94	24.56	13158.51
Maximum	68.60	31.47	40.10	17782.00

2.3 Machine Learning Methods

Recent work from Angra and Ahuja (2017) explored some of the new effective methods of using machine learning in data analytics, modelling and algorithms. Great results can be observed when machine learning models are used in pattern and feature recognition through learning trends and biases in data sets. Differently from traditional computational methods, machine learning methods are able to use the training data provided to suitably adjust its parameters to make better future predictions.

Many branches of science are exploring and experimenting with the use of machine learning on research projects. Tarca *et al.* (2007) thoroughly explored the use of some methods of machine learning in the field of biology, for instance, methods like support vector machines and clustering are discussed. An important consideration made by Tarca *et al.* (2007) is that, even though modern biology can benefit from advancements made in the area of machine learning, one should be cautious about choosing a specific machine learning approach over another, sometimes the results could be misleading and interpretation can become challenging. This consideration should be extrapolated to all fields of science in which machine learning methods could be applied, bad analysis of the results can, and will, lead to misleading conclusions.

2.3.1 The Decision Tree and Related Methods

A decision tree is a machine learning model that can be used to classify and predict data (De Ville, 2013). A root node, branches, internal nodes, and leaf nodes all make up the tree structure of the model. Decision trees grow through the addition of question nodes incrementally. The tree uses a series of, ideally homogeneous, binary partitions on specific variables to separate the predictor space. The base node contains all predictor space. The terminal nodes are the nodes that aren't split and make up the last partition of the predictor space. Depending on the value of one of the predictor variables, each non-terminal node divides into two descendant nodes, one on the left and one on the right (Kingsford and Salzberg, 2008).

This model may be accurate for the training data, but, when the test data is used to generalize to data not found in the training, the model is quite likely to fail (James *et al.*, 2013). This concept is called "overfitting". To address the issue of overfitting trees, in this work two other methods based on decision trees will be employed. Those methods are more accurate and have less chance of overfitting, being the first one called "Random Forest" and the second one "Extreme Gradient Boosting".

2.3.2 Random Forest

The random forest (RF) algorithm, introduced by Breiman (2001), is a machine learning model that can be used to classify and predict data based on a decision tree model. The main idea of this model is to create a combination of decision randomly generated trees. Due to the amount of randomly generated trees, the model has less chance of overfitting, since the averaging of uncorrelated trees lowers the overall variance and prediction error.

A random forest method can provide good results without the need of hyperparameter tuning (Fernández-Delgado *et al.*, 2014), although, it's possible to achieve better performance profiting from hyperparameter tuning (Probst *et al.*, 2019). Three main hyperparameters influence the most on the performance of a random forest model: The number of decision trees, the minimum number of samples to be split and the maximum depth of the tree. (Sun *et al.*, 2020).

The prediction outcome on a Random Forest model is the result of several individual decision tree predictions averaged to produce a final prediction outcome. In a regression model, in order to get the final prediction, the average result of all individual predictions is calculated (Palmer *et al.*, 2007). A scheme of how this model work, is illustrated in Figure 2.

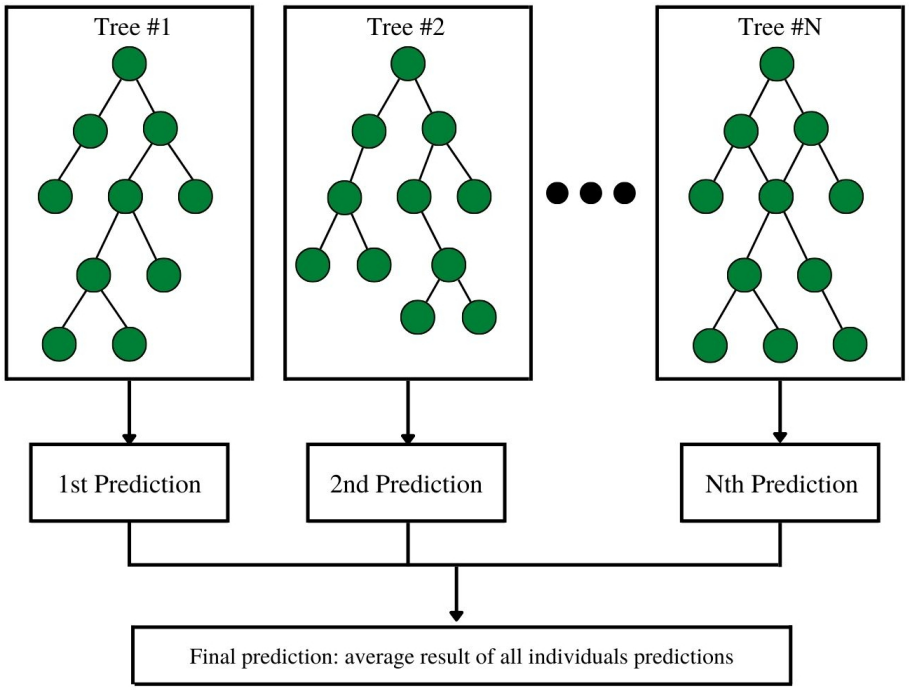

Figure 2. Architecture of a generic Random Forest Model

2.3.3 Extreme Gradient Boosting

The other tree-based method that will be used in this work is the "gradient boosting" method. It consists of a set of weak learning models, such as decision trees, to make predictions, the same concept used in the Random Forest method. However, in this instance, all of the decision trees are generated using information from previously grown trees generated so far (James *et al.*, 2013), unlike the Random Forest model, where all trees are randomly generated. The final prediction value is obtained using the same method as the Random Forest model.

In this work, the extreme gradient boosting method (XGB) is used. The main difference between both methods is that XGB incorporates a regularized model to prevent overfitting, along with improved gradient approximation. Accordingly to Bentéjac *et al.* (2021), the most important hyperparameters for this model are the learning rate, the number of trees, the maximum depth of the tree, the minimum loss reduction required to make a further partition on a leaf node of the tree and the subsampling rate.

2.3.4 Artificial Neural Networks

Finally, the last method of machine learning used in this work is "Artificial Neural Networks" (ANN), precisely, a feedforward neural network. An artificial neural network, differently from Random Forests and Gradient Boosting, isn't based on decision trees. Even though this kind of ANN is the most straightforward kind of neural network, it's more than enough for the type of modeling presented in this work. According to Nielsen (2015), a feedforward neural network is composed of 2 basic layers, input and output layers, combined with an arbitrary number of hidden layers. The input layer is usually represented as the leftmost layer and contains the data fed to the neural network. Similarly, the output layer is represented as the rightmost layer and it contains the data we want to get from the neural network. The intermediary layers are called hidden layers. The number of hidden layers can vary a lot depending on many factors, being almost impossible to clearly define a good "rule of thumb" on how many hidden layers to use. In the hidden layers, all the calculations and adjustments on all of the neurons' weights and biases will be made, working towards an objective, for example, minimizing the loss function. A generic example of a feedforward neural network is presented in Figure 3.

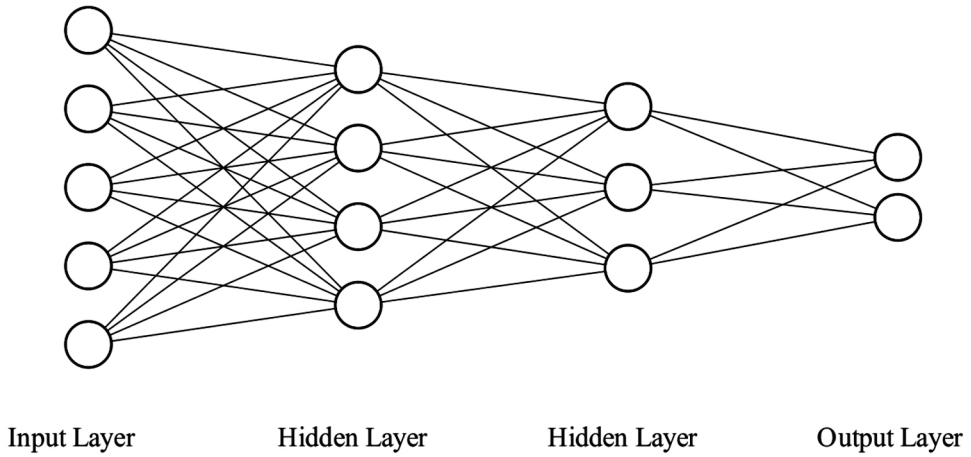

Figure 3. Example of a generic feedforward neural network architecture.

To understand how feedforward neural networks operate, it's crucial to understand what occurs inside a single neuron. First of all, the neuron will receive the input signals, being those inputs the ones in the input layer or being those inputs the outputs of a previous layer, and each signal will have a numerical value associated with it.

Within the neuron, 2 main components that will be responsible for changing all of the input signals received, those components are called "activation function" and "bias". The first component will be a function that will change the value of the weighted sum of the inputs. Many different functions can be used as an activation function, such as: tanh, sigmoid, reLU, linear, etc. The "bias" parameter is used to slightly shift the activation function left or right, serving as a small numerical adjustment in the output. Sometimes it can also be seen as a numerical threshold for the activation, or not, of the neuron. These are the parameters adjusted simultaneously, on all of the network's neurons, working towards the goal of minimizing the loss function. Figure 4 illustrates the basic structure of a simple neuron.

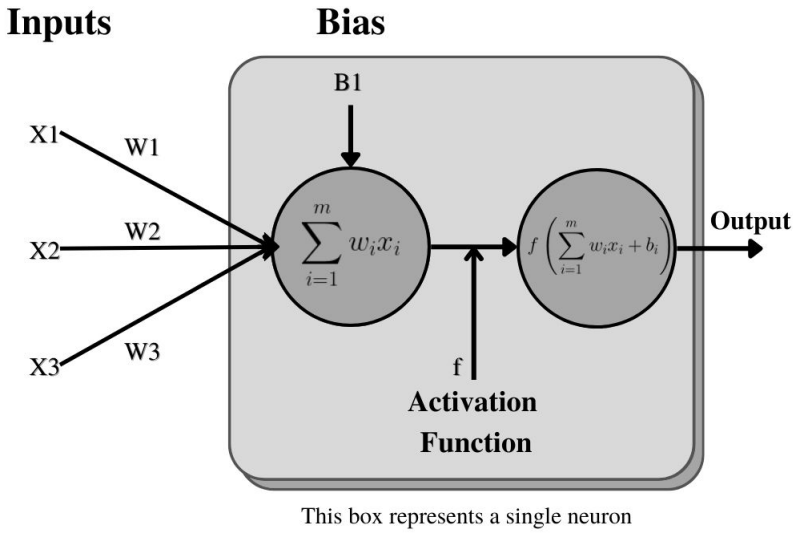

Figure 4. Basic structure of a simple neuron.

2.4 Statistical metrics

To check the validity of all the correlations established, two major performance metrics were chosen to verify the goodness of fit. The first metric used is the "Correlation Coefficient", usually called "Pearson's R". This metric quantitatively measures how deeply different variables are related, as well as reflects the noisiness and direction of linear relationships (Haslwanter, 2016).

Finally, the last metric used was the "Mean Squared Error". This metric is fairly simple to understand when compared to the other two. The "MSE" is, simply put, the expected squared difference between two estimators (Montgomery and Runger, 2003).

These metrics are calculated according to the equations 1 and 2, where y is the value estimated by the model, $\hat{y}$ is the experimental value, $m_{\hat{y}}$ is the average of the experimental values and m_y is the average of the estimated values.

$$R = \frac{\sum_{i=1}^{n}(y_i - m_y)(\hat{y}_i - m_{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - m_y)^2}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - m_{\hat{y}})^2}} \tag{1}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2}$$

3. RESULTS AND DISCUSSION

All machine learning models were implemented in *Python3*, using different libraries. The artificial neural network (ANN) model was developed using a deep learning library called *TensorFlow* running along with a *Keras* API. The random forest model was developed using the *Scikit-Learn* library, the *XGBoost* library was used to develop the extreme gradient boost (XGB) model and, finally, the linear regression model was developed using *Scipy*.

To avoid overfitting, the dataset was divided between training and testing, being 70% of the dataset only for training and 30% only for testing. This division was performed randomly. All of the models were evaluated and compared according to the statistical parameters mentioned in the previous section, R and MSE.

Table 3 illustrates the performance of all developed neural network models. All models have a learning rate of 0.001, with just one hidden layer and the output layer activation function was set to linear. For comparison purposes, the number of neurons in each network was changed, ranging from 5 to 15. The activation function on the hidden layer was changed between tanh and reLU, for the sake of comparison. And as a result, the best neural network architecture, that is, the one that presents the best statistical parameters, for the interpretation of the data has 15 neurons and its activation function is tanh.

Table 3. Performance analysis of all ANN models studied.

Neurons	Activation Function	R	MSE
5	tanh	0.9134	0.0267
5	reLU	0.9137	0.0269
10	tanh	0.9186	0.0251
10	reLU	0.8937	0.0327
15	tanh	0.9191	0.0249
15	reLU	0.8981	0.0314

The XGB and the RF models had their hyperparameters optimized using a Bayesian optimization, through the *scikit-optimize* library. Table 4 summarizes the process of optimization. Second and fourth columns describe the search range, while third and fifth columns display the best value found.

Table 4. Optimisation of the XGB and RF hyperparameters.

Hyperparameters	XGB Range	XGB Value	RF Range	RF Value
Learning rate	0.03 - 0.3	0.1845	-	-
Number of decision trees	10 - 500	312	10 - 500	493
Maximum depth of each tree	2 - 10	2	2 - 10	7
Minimum number of samples needed by the leaf	2 - 10	3	2 - 10	6
Subsample ratio	0.0 - 0.5	0.3	-	-
Minimum number of samples to be split	-	-	0.0 - 0.5	0.136

The best result for the XGB model was with a learning rate of 0.1845, 312 decision trees, with a maximum depth of 2 for each tree, the minimum number of observations needed in a child node of 3 and the subsample ratio of 0.3. The Random Forest model used 493 decision trees, with a maximum depth of 7, a minimum number of samples that the leaf needs to have equal to 6 and the minimum number of samples to be split equally to 0.136.

It's seen from the results presented in Table 5, that each of the machine learning models has good prediction accuracy and generalization capability by achieving R > 0.89 and MSE < 0.035. However, the ANN model exhibited a relatively high performance to predict the HHV from the proximate analysis, reaching R = 0.9191 and MSE = 0.0249. The Linear Regression can be introduced as the second more accurate model, showing R = 0.9153 and MSE = 0.0265, which can be explained by the possible strong linear correlation between the data seen in Figure 1. As it can be seen, the two tree-based models, Random Forest and XGB, did not perform as good as both previous models, barely reaching R = 0.90 and having mean squared errors greater than 0.03.

Table 5. Performance analysis of all machine learning models.

Model	R	MSE
Neural Network	0.9191	0.0249
Random Forest	0.9018	0.0300
Extreme Gradient Boost	0.8931	0.0353
Linear Regression	0.9153	0.0265

To gain a better insight into the prediction success of the model, Figure 5 shows three different scatter plots for training, test and the entirety of data, respectively. These scatter plots show how correlated the actual and the predicted values by the ANN are. It's possible to see that the values predicted by the ANN model match the experimental values satisfactorily.

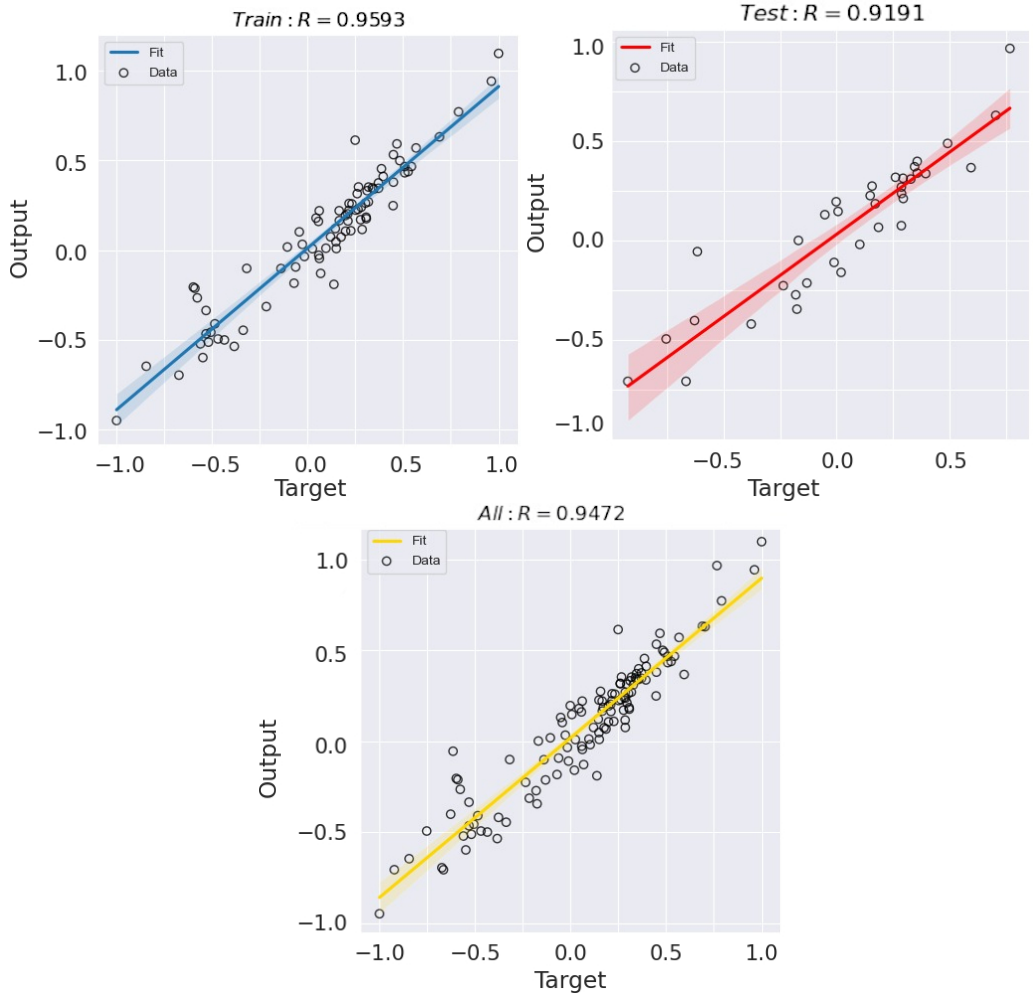

Figure 5. Graphical comparison of experimental and estimated higher heating value with ANN model.

The linear regression model was obtained and had Eq. 3 as result, being A the ash content (%), VM volatile matter (%) and FC fixed carbon (%), as mass percentage in dry basis. The higher heating value (HHV) has units of kJ/kg. Even though the linear regression did not show the best statistical metrics (R = 0.9153 and MSE = 0.0265), it's important to note that the linear model expressed good results considering it is a relatively simple model. The simplicity of the model not only helps with ease of application but also reduces the chance of overfitting.

$$HHV = -15.876 \cdot (A) + 225.928 \cdot (VM) + 360.245 \cdot (FC) \tag{3}$$

To summarize, it was found that all models studied have a good ability to predict the HHV of Candiota coal since all models achieved an R > 0.85. The possible reason why the Artificial Neural Network model displayed better results when compared to the other models, could be explained by data normalization. Data normalization was performed on all models because the ANN requires data to be normalized. So, to maintain consistency when comparing all four models, data normalization has been applied in all of them. This could be the reason why the ANN showed a slight performance increase in comparison to the other models.

It can be seen that both the size and consistency of the dataset used to train the machine learning models had a significant impact on the models' performance. Due to the small number of samples, it was not possible to develop more complex and well structured models. A complex model, built using this same dataset, would unavoidably incur in overfitting. The consistency of the samples also played a major role on the performance results. It can be seen, in Figure 1 that the datapoints, notably observed in the volatile matter plot, suffer from a moderate degree of variance due to data collected from different seams mixed together. This observed variance most likely had a negative impact on all of the models' performance.

As previously mentioned, it is a hard task to compare the models developed in this work since no other correlations using the proximate analysis to find the HHV were developed specifically for Candiota coal. That being said, two similar works were developed by Akkaya (2009) and Tan *et al.* (2015). Akkaya (2009) obtained an accurate model (R^2 of 0.97) using 143 Turkish low rank coal samples for a non linear regression using four variables: moisture, fixed carbon, ash and volatile matter. Tan *et al.* (2015) used a machine learning method called support vector regression to develop a non linear model for both 167 Chinese coal samples and 4540 U.S. coal samples (R of 0.986 and 0.996, respectively). It is possible to see that the models developed on both of these works outperform the models developed here. The main possible cause could be the size of the datasets used to develop those correlations and the greater number of variables used on the models. That being said, that does not mean that those models could necessarily be used to accurately predict the HHV for Candiota coal, since this kind of correlation is very dependent on the geographic region of the coal. Therefore, more accurate correlations between the higher heating value and a coal's proximate analysis for Candiota coal could be developed. This would, most likely, require a larger dataset, taking into account that specific sample's seam.

4. CONCLUSION

Establishing a strong relationship between the variables obtained from the proximate analysis and the higher heating value of coal sampled using regression methods of machine learning posed a great challenge. The best performing model was developed using an artificial neural network, having statistical metrics of R = 0.919 and MSE = 0.0249, this was the only model able to statistically perform better than a linear regression. A major contributing factor to the somewhat unrefined results, when compared to other correlations established for different coals, can be pointed out as the lack of a large database for Candiota coal. Another factor that may have had some negative impact on the performance of the machine learning methods was the heterogeneity of coal composition data since the dataset contained data of coal samples extracted from different seams mixed. A possible workaround to improving performance could be adding the seam as a categorical input for the neural network. Even though this is theoretically possible, not nearly enough data was provided for this application, in fact, this would make the amount of data needed for the analysis increase significantly

The results, though, provided a good insight into how machine learning models can be effective tools, given the right amount of data and fine tuning, to establish specific correlations between variables that have no apparent phenomenological correlations. Adding to that, machine learning models can be really powerful tools when dealing with relationships between properties that cannot be generalized due to some kind of specificity on the subject of study, in this case, the coal mined in Candiota coalfield.

5. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the data provided by Companhia Riograndense de Mineração (CRM). Authors also acknowledge the CNPq and the FAPERGS for awarding the scholarships of the first and second authors, respectively.

6. REFERENCES

Akhtar, J., Sheikh, N. and Munir, S., 2017. "Linear regression-based correlations for estimation of high heating values of Pakistani lignite coals". *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, Vol. 39, No. 10, pp. 1063–1070. ISSN 1556-7036, 1556-7230. doi:10.1080/15567036.2017.1289283.

Akkaya, A.V., 2009. "Proximate analysis based multiple regression models for higher heating value estimation of low rank coals". *Fuel Processing Technology*, Vol. 90, No. 2, pp. 165–170.

Angra, S. and Ahuja, S., 2017. "Machine learning and its applications: A review". In *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*. IEEE, pp. 57–60.

Bentéjac, C., Csörgő, A. and Martínez-Muñoz, G., 2021. "A comparative analysis of gradient boosting algorithms". *Artificial Intelligence Review*, Vol. 54, No. 3, pp. 1937–1967.

Breiman, L., 2001. "Random forests". *Machine learning*, Vol. 45, No. 1, pp. 5–32.

Companhia Riograndense de Mineração - CRM, 2008. "Jazida de candiota". Companhia Riograndense de Mineração. 06 Jun. 2022 <http://www2.crm.rs.gov.br/>.

De Ville, B., 2013. "Decision trees". *Wiley Interdisciplinary Reviews: Computational Statistics*, Vol. 5, No. 6, pp. 448–455.

Fernández-Delgado, M., Cernadas, E., Barro, S. and Amorim, D., 2014. "Do we need hundreds of classifiers to solve real world classification problems?" *The journal of machine learning research*, Vol. 15, No. 1, pp. 3133–3181.

Goodfellow, I., Bengio, Y. and Courville, A., 2016. *Deep Learning*. MIT Press. `http://www.deeplearningbook.org`.

Haslwanter, T., 2016. *An Introduction to Statistics with Python: With Applications in the Life Sciences*. Statistics and Computing. Springer International Publishing, 1st edition. ISBN 978-3-319-28315-9,978-3-319-28316-6.

James, G., Witten, D., Hastie, T. and Tibshirani, R., 2013. *An introduction to statistical learning*, Vol. 112. Springer.

Kalkreuth, W., Holz, M., Kern, M., Machado, G., Mexias, A., Silva, M., Willett, J., Finkelman, R. and Burger, H., 2006. "Petrology and chemistry of Permian coals from the Paraná Basin: 1. Santa Terezinha, Leão-Butiá and Candiota Coalfields, Rio Grande do Sul, Brazil". *International Journal of Coal Geology*, Vol. 68, No. 1-2, pp. 79–116. ISSN 01665162. doi:10.1016/j.coal.2005.10.006.

Kalkreuth, W., Lunkes, M., Oliveira, J., Ghiggi, M., Osório, E., Souza, K., Sampaio, C. and Hidalgo, G., 2013. "The lower and upper coal seams of the Candiota Coalfield, Brazil — Geological setting, petrological and chemical characterization, and studies on reactivity and beneficiation related to their combustion potential". *International Journal of Coal Geology*, Vol. 111, pp. 53–66. ISSN 01665162. doi:10.1016/j.coal.2012.09.013.

Kingsford, C. and Salzberg, S.L., 2008. "What are decision trees?" *Nature biotechnology*, Vol. 26, No. 9, pp. 1011–1013.

Montgomery, D.C. and Runger, G.C., 2003. *Applied statistics and probability for engineers*. Wiley, New York, 3rd edition. ISBN 978-0-471-20454-1.

Nielsen, M.A., 2015. *Neural networks and deep learning*, Vol. 25. Determination press San Francisco, CA, USA.

O'Keefe, J.M., Bechtel, A., Christanis, K., Dai, S., DiMichele, W.A., Eble, C.F., Esterle, J.S., Mastalerz, M., Raymond, A.L., Valentim, B.V. *et al.*, 2013. "On the fundamental difference between coal rank and coal type". *International Journal of Coal Geology*, Vol. 118, pp. 58–87.

Palmer, D.S., O'Boyle, N.M., Glen, R.C. and Mitchell, J.B., 2007. "Random forest models to predict aqueous solubility". *Journal of chemical information and modeling*, Vol. 47, No. 1, pp. 150–158.

Probst, P., Wright, M.N. and Boulesteix, A.L., 2019. "Hyperparameters and tuning strategies for random forest". *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, Vol. 9, No. 3, p. e1301.

Speight, J.G., 2015. *Handbook of coal analysis*. John Wiley & Sons.

Stracher, G.B., Prakash, A. and Sokol, E.V., 2010. *Coal and peat fires: a global perspective: volume 1: coal-geology and combustion*, Vol. 1. Elsevier.

Sun, D., Wen, H., Wang, D. and Xu, J., 2020. "A random forest model of landslide susceptibility mapping based on hyperparameter optimization using bayes algorithm". *Geomorphology*, Vol. 362, p. 107201.

Tan, P., Zhang, C., Xia, J., Fang, Q.Y. and Chen, G., 2015. "Estimation of higher heating value of coal based on proximate analysis using support vector regression". *Fuel Processing Technology*, Vol. 138, pp. 298–304.

Tarca, A.L., Carey, V.J., Chen, X.w., Romero, R. and Drăghici, S., 2007. "Machine learning and its applications to biology". *PLoS computational biology*, Vol. 3, No. 6, p. e116.

UN-ECE, 1998. "International classification of in-seam coals". *UN ECE: Geneva, Switzerland*.

Ward, C.R. and Suárez-Ruiz, I., 2008. "Introduction to applied coal petrology". *Applied coal petrology*, pp. 1–18.

7. RESPONSIBILITY NOTICE

The authors are solely responsible for the printed material included in this paper.