Predicting Moisture Content on Wood Using Machine Learning Classification Methods

Vítor Mendes Magalhães[a], Giancarlo Lucca[b], Alessandro de Lima Bicho[c] and Eduardo N. Borges[d]
Centro de Ciências Computacionais, Universidade Federal do Rio Grande, FURG, Brazil

Keywords: Moisture Content, Wood, Intelligent Systems, Machine Learning, Artificial Intelligence.

Abstract: The growing demand for wood in several industry segments and for its economical value increased illegal deforestation in several countries. As a direct consequence, climate changes across the planet have been aggravated, which further increases the prominence and concern about the issue of deforestation. So that these potentially catastrophic effects can be mitigated, it is necessary to better use wood in production processes. In this sense, a key point is the variation of the moisture content of the wood as a function of storage time, since, as the wood logs are stored outdoors, they gradually begin to lose water. Dry wood usually cracks, which makes most of its use unfeasible – depending on the purpose – which can even lead to the disposal of the log. Considering that there is a direct relationship between moisture content and wood weight, this work aims to develop different possible solutions for this problem using explainable machine learning methods, contributing to the effectiveness in controlling the variation in moisture content and, consequently, to a better use in the production processes in which wood is used as a raw material.

1 INTRODUCTION

Different countries has increasingly established its position as an exporter of natural products – as for example, Brazil. Precisely, the Brazilian relation with the Asian market have increased. With the growing demand from Chinese industry, exports have been gradually required, and this has had a great impact both on the organization of Brazilian agriculture and on its technological configuration (Vieira et al., 2019).

Brazil has been studying and forecasting an increase in demand for wood since the 1960s, when tax incentives were created for farmers who produce wood by planting fast-growing species, such as *Pinus* and *Eucalyptus* (Kengen, 2001).

As much as the advances are continuous, historically Brazil has difficulties about promoting the expansion of planted forests while controlling deforestation. When dealing with farmers who are not used to planting trees for this purpose, some specific characteristics of this market can greatly reduce its attractiveness. The fact that the financial return is only realized after many years can be cited as the main cause of the lack of attractiveness.

Deforestation is a growing concern, especially for developing countries. It has global repercussions, as forest losses can directly imply changes in the water balance, in the carbon cycles and obviously in the supply of wood (Allen and Barnes, 1985). Even the United Nations (UN) consider this issue as one of the 17 goals to a sustainable development[1].

Regardless of the destination that will be given to the wood, many products that have wood as raw material go through the same stage: the storage of wood logs in piles. But as the storage time passes, the moisture content on wood will reduce (Rezende et al., 2010). Such changes directly alter all mechanical properties of wood.

While the wood logs are stored in piles and exposed to weather conditions, their weight decreases due to moisture loss (Tomczak et al., 2018). For this reason, the storage time of the logs in the piles is decisive (Lima et al., 2017; Júnior and Alves, 2019). So,

[a] https://orcid.org/0000-0003-3588-9930
[b] https://orcid.org/0000-0002-3776-0260
[c] https://orcid.org/0000-0002-6572-1496
[d] https://orcid.org/0000-0003-1595-7676

[1] For more informations see – https://www.un.org/sustainabledevelopment/sustainable-development-goals/

Magalhães, V., Lucca, G., Bicho, A. and Borges, E.
Predicting Moisture Content on Wood Using Machine Learning Classification Methods.
DOI: 10.5220/0011988600003467
In *Proceedings of the 25th International Conference on Enterprise Information Systems (ICEIS 2023) - Volume 1*, pages 607-614
ISBN: 978-989-758-648-4; ISSN: 2184-4992

the moisture content is considered a key factor in the storage of wood, and ideally it should be kept within the standards (Estuqui Filho, 2006), because the storage time can lead to losses in the production process – precisely because of cracks – caused by moisture loss.

Companies in the sector, seeking to reduce and even avoid the cracks problem, use some devices. One of the most used devices is the *anti-split*, which basically consists of two metal plates installed at the ends of the logs (cross sections).

It was not possible to find machine learning-based methods to predict the variation of moisture content in wood logs, as a function of the weight of the logs. In this sense, there is a study that presents a review, with several other methods published in recent years, which use other variables to solve the current problem (Magalhães et al., 2022).

So, the aim of this work is to develop a Artificial Intelligence-based model (Russell, 2010) – specifically using machine learning classification methods (Tan et al., 2005; Tavana et al., 2022) – to predict the moisture content on wood. Precisely, we have consider the application of different methods that are able to produce explainable models, which can be latter used to better understanding the problem.

This paper is organized as follows. First, the theoretical foundation of the present work is presented in Section 2. Then, the methodology, describing the considered dataset and pre-processing is shown in Section 3. After that, in Section 4, the results are discussed. At the end, the main conclusions are drawn.

2 THEORETICAL FOUNDATION

This section provide the concepts related with the paper. It starts presenting the concepts of machine learning and the considered methods.

2.1 Machine Learning

The machine learning process can use different methods to solve problems. Usually, it is emphasized that there is no single approach that best solves all problems. Therefore, it is important to incorporate the specific knowledge of the problem into the behavior of the algorithm(Yazdi et al., 2018; Khorhidpoor et al., 2023), as well as to understand the limitations of the algorithms, preferably using methodologies that allow evaluating the concepts induced by them in the resolution of certain problem (Mahesh, 2020).

When dealing with supervised learning, verifying the dataset and its relationship with the problem that must be solved, it is necessary to analyze the target attribute from two different perspectives: classification and regression problems (Harrington, 2012). As some classes will be created to represent the weight loss intervals of wood logs as a function of moisture content variation (see subsection 3.1.3), in addition to being a supervised machine learning problem, should be treated as a classification problem.

2.2 Considered Algorithms

There are different approaches to deal with classification problems. Decision Trees – DT (Freund and Mason, 1999), Support Vector Machines – SVM (Steinwart and Christmann, 2008), Fuzzy Rule-Based Classification Systems – FRBCS (Cordón et al., 1999) and Artificial Neural Networks – ANN (Yegnanarayana, 2009) are a few approaches. It is necessary to point out that each one have a large set of related algorithms. In what follows, the considered algorithms used in this study are introduced.

It is important to observe that each algorithm produce an interpretable model – that can be used to better understand the decision made in the prediction of new examples.

FURIA: Fuzzy Unordered Rule Induction Algorithm – FURIA (Hühn and Hüllermeier, 2009) is an algorithm that consider the IREP (*Incremental Reduced Error Pruning*) to generate the rules, improving the performance in comparison to the use of a default rule.

RIPPER: Repeated Incremental Pruning to Produce Error Reduction – RIPPER (Cohen, 1995) is one of the most used algorithms for rule induction. It orders, in an ascending way, the classes involved in the problem according to their frequency in the training set, being suitable for the development of models that deal with unbalanced datasets.

C4.5: Based on decision trees which is able to deal with continuous values, unavailable values, prune the trees and derive rules from that, C4.5 (Quinlan, 1993) aims to generate a classifier model presenting two different states during the process: a leaf and a decision node. Based on the attribute under analysis, it may result in a branch, or a subtree, for each value found in the base.

Random Forest: Being a classifier formed by a set of classification trees, each constructed from a random sampling of the original training set, Random Forest – RF (Breiman, 2001) is a algorithm that obtains the forests through bootstrapping aggregating, a method used to generate multiple versions of a predictor, that are built re-sampling the original set. The classification of a feature vector is done by voting.

3 METHODOLOGY

The methodology adopted in this study is described in this section. We start by the issues related with the dataset, such as data integration and exploratory analysis. Later, the statistical tests are discussed and the parameters used by the algorithms are shown.

3.1 Dataset

The data used for the construction of specific prediction models have different origins. For this reason, it becomes very important to describe the integration process of different databases, resulting in the final dataset considered in this research.

When it comes to data composition, it can be categorized into two distinct groups: *operational data* and *meteorological data*.

- The **operational data** are the specific storage data of the wood logs, such as information about the piles, dates and weights; and biological-forestry data of the logs, such as the type of wood, species, length, diameter class – as required by the international market, and the presence (or not) of bark.
- The **meteorological data** are the atmospheric or climatic data that have relationship with the variation of the moisture content in the wood. Such data are extremely important for the construction of prediction models because it is known that the process of variation in moisture content is also based on specific meteorological data.

These two groups of data were obtained from different sources. For this reason, its categorization is even more important. Next, the ways of obtaining data will be detailed, as well as their description.

3.1.1 Obtaining Data

The *operational data* were provided by a company whose purpose is to buy Brazilian wood and then export to the European and Asian markets, especially to China. The referred company's log storage yards are located in the city of Rio Grande, state of Rio Grande do Sul, in the south of Brazil.

As the exact location of the storage yards is known, it was possible to obtain all historical meteorological data available by the specific measuring station through the National Institute of Meteorology (INMET)[2] – officially linked to the Ministry of Agriculture, Livestock and Supply of Brazil (MAPA)[3]. Then, in INMET's own system on the internet[4], the automatic measurement station closest to the storage yards was searched. Precisely, the data were obtained from the A802[5] measuring station, located at coordinates 32º04'43.7" S 52º10'03.8" W.

For each existing pile of logs in the database of the referred company's storage yards, was obtained the first date on which any input occurred and also the last date on which any in/out happened. With this period in hand, historical meteorological data for the same location were searched in the INMET database – meteorological data recorded daily and automatically by measurement stations.

As the equipment of the automatic meteorological stations can have problems, some dates were found without any measurement. For these instances, the arithmetic mean between the two closest dates – before and after – in which there is measurement was defined.

3.1.2 Description of the Variables

The generated dataset, then, resulted in a total of 759 instances and 23 different attributes, divided into nominal, ordinal and categorical types. It contains all the data that can be extracted from the different sources, categorized into *operational data* and *meteorological data*, as explained above.

The list of all the attributes of the dataset is shown in Table 1. For each one of them, the data type is presented, as well as its minimum, maximum and mean (mode for categorical and median for ordinal), as well as its description for a better understanding.

3.1.3 Pre-Processing Data

Starting the data pre-processing step, a new attribute was created in the dataset, named *PERCENTAGE*. It represents the percentage of wood weight loss (reversals) in relation to the sum of wood inputs in each pile, and can be represented through the equation $P = (E \times 100) \times T^{-1}$. Where P is the percentage of weight loss, E represents the reversed weight of each pile, and T represents the sum of the input weights of wood in each pile.

After creating the attribute, instances representing piles from which there was no record of output were removed. In theory, either such instances of piles continued to receive loads of wood, or they were not yet

[2]https://portal. inmet.gov.br/
[3]https://www.gov.br/agricultura/pt-br
[4]https://bdmep.inmet.gov.br/
[5]https://tempo.inmet.gov.br/TabelaEstacoes/A802

Table 1: Details of the attributes present in the final dataset.

Attribute	Type	Min	Max	Average/Mode	Description
pile	categorical	na	na	na	id of the wood log storage stack
initial_date	date	na	na	29/03/2022	start date of depositing logs in the pile
final_date	date	na	na	24/05/2021	end date of deposits and/or withdrawals of logs from the pile
product	categorical	na	na	eucalyptus	type of wood stored in the pile (*pinus* or *eucalipto*)
species	categorical	na	na	grandis	species of wood stored in the pile
diameter	ordinal	8 a 12	40 a 60	20 a 30	cross-section diameter class, measured by the thinnest end of the wood logs in the pile
length	ordinal	3,1	6	5,2	length class (size, in meters) of wood logs
bark	categorical	na	na	yes	whether the wood logs in the pile have bark or not
temp_min	continuous	1,9	24	5,4	minimum temperature (in °c) of the period
temp_max	continuous	12,7	39,4	28,5	maximum temperature (in °c) of the period
temp_mean	continuous	7,73	37,7	17,2	average temperature (in °c) of the period
temp_po	continuous	4,2	29,4	12,6	average temperature (in °c) of the period in which the water present in the ambient air changed to a liquid state
precipitation	continuous	0	1156	95,4	total precipitation (in millimeters) for the period
atmp_mean	continuous	1.005	1.535,4	1.016,4	average atmospheric pressure (in hpa) of the period
ru_min	continuous	21	91	32	minimum relative humidity of the air (in %) of the period
ru_med	continuous	59,5	117,5	76,4	average relative humidity of the air (in %) of the period
wind_mean	continuous	1,1	5,4	3	average wind speed (in m/s) over the period
raj_max_mean	continuous	3,7	19,3	10,3	average maximum wind gust speed (in m/s) for the period
qtd_days	discrete	0	425	47	number of days in the period
tot_in	continuous	940	8.128.820	216.740	total entries (in kg) of logs in the pile
tot_out	continuous	0	6.628.260	168.480	total outputs (in kg) of logs from the pile
difference	continuous	-169.280	1.500.560	18.990	difference (in kg) between total inputs and total outputs
reversals	continuous	0	1.500.560	9.220	amount reversed (in kg) of the pile – in theory, this attribute should be directly related to water loss

finalized at the time of obtaining the *operational data*. They may also represent data entry errors by the operational sector of the company that provided the data.

Instances that had no reversed amount were also removed, that is, instances representing piles that, in theory, had no weight loss. Such instances need to be removed based on the specialist's knowledge: in natural drying, under the conditions of the captured data, there will always be a loss of weight.

After removing these instances, the feature selection process was started. In real-world situations, in which data are not available in the ideal format to start the knowledge discovery process – and are often obtained from different sources, it is necessary to use of tools that make data mining more effective.

For the creation of classification models, it is necessary to create a new attribute, which is called *LOSS_INTERVAL*. This attribute will represent the weight loss percentage intervals of each pile in relation to the total inputs. We present it in Table 2, where is possible to observe the total number of instances resulting in each weight loss interval.

Table 2: Breakdown of weight loss intervals.

Name of the interval	**Interval**	**Samples**
A	$\geq 0\%$ and $\leq 10\%$	478
B	$\geq 10\%$ and $\leq 20\%$	173
C	$\geq 20\%$ and $\leq 30\%$	58
D	$\geq 30\%$	17

After analyzing the data in the Table 2 (*samples* column), it can be observed that the largest number of instances of the dataset is represented in classes A and B, that is, the vast majority of the piles have a loss of weight, in percentage, less than 20%. It is necessary to verify the integrity of this information.

One possible way is to compare the weight loss with the number of days of storage in the piles. After all, the longer is the storage time, the greater will be the weight loss – until the variation is close to zero.

When it comes to the loss of moisture content in wood stored outdoors, 20% is a small percentage. It can be inferred that the observed piles, due to the average number of days of storage (almost 65 days), had already been cut longer – information discovered from descriptive data analysis and which is important for the subsequent analysis of predictive models.

With this, it can be understood that the final datasets intended for the development of the classification models is ready to be submitted to the next step of this research, the *data mining*, being composed by 726 instances.

3.2 Parameter Set-up

In this paper, 5 different algorithms are considered to tackle the proposed classification problem. These algorithms were applied through the WEKA software[6].

In Table 3, for each algorithm, the used parameters are shown. We highlight that these are the default ones used in the tool.

Table 3: Set-up of the hyperparameters used by the machine learning algorithms.

Algorithm	Hyperparameters	
FURIA	T-norm: Product batchSize: 100 checkErroRate: True debug: False doNotCheckCapabilities: False folds: 3	minNo: 2.0 numDecimalPlaces: 2 optimizations: 2 seed: 1 uncovAction: Rule streching
Ripper	batchSize: 100 checkErroRate: True debug: False doNotCheckCapabilities: False folds: 3	minNo: 2.0 numDecimalPlaces: 2 optimizations: 2 seed: 1 usePruning: True
C4.5	batchSize: 100 binarySplits: False collapseTree: True confidenceFactor: 0.25 debug: False doNotCheckCapabilities: False doNotMakeSplit: False minNumObj: 2 numDecimalPlaces: 2	numFolds: 3 reducedErrorPruning: False saveInstanceData: False seed: 1 subtreeRaising: True unpruned: False useLaplace: False useMDLcorrection: True
RF	bagSizePercent: 100 batchSize: 100 BreakTiesRamdomly: False calcOutOfBag: False ComputeAttributeImportance: False debug: False doNotCheckCapabilities: False maxDepth: 0	NumDecimalPlaces: 2 numExecutionSlots: 1 numFeatures: 0 numIterations: 100 OutputOutOfBagComplexity: False PrintClassifiers: False Seed: 1 storeOutOfBagPredictions: False
Baseline	BatchSize: 100 debug: False doNotCheckCapabilities: False numDecimalPlaces: 2	

In this study, we validate the models considering a hold-out approach. That is, the original dataset is splitted into different partitions of *training* and *test*. To avoid a split that facilitate the model training, we have considered 5 different runs with different values (and amount of data) in each one, setting a seed[7]. The relation of seeds and the considered runs are: 1 (Run 1), 1234 (Run 2); 500 (Run 3); 98765 (Run 4) and 999999 (Run 5).

4 EXPERIMENTAL RESULTS

In this section, the obtained results are shown. Precisely, the results are provided in Table 4, where for each hold-out configuration we provide the obtained

[6]For more information about this software, visit – https://www.cs.waikato.ac.nz/ml/weka/.

[7]A seed is a number used in the pseudo random value generator. Observe that it is possible to set a value to this parameter to guarantee the reproducibility of the experiment.

accuracy for different runs and methods. we point out that the baseline is considered as the majority class.

In order to ease the comprehension of the obtained results, we highlight for each execution with **boldface** the highest accuracy and underline the lowest. Similarly, in order to provide a general analysis among all the experiment, we check with $^{\uparrow}$ the largest general accuracy and with $^{\downarrow}$ the lowest.

Starting with a general analysis of the obtained results, it is noticeable that the FURIA method achieved the largest accuracy mean in 3 different runs. The RF method presented one largest accuracy run for the last run in the hold-out 75-25. Also, it is interesting to notice that the Baseline was not outperformed for any method for the run 4 in the hold-out 90-10. This behavior is probably due to the generalization caused by the division of the data.

Taking into account the cases where the approaches provided the lowest general accuracy, the Baseline is outperformed by 4 out the 5 runs. It can be considered as an expected behavior since this is a simple approach that indicate if the learned models can be considered as satisfactory. However, it is necessary observe that for the RF method, in one specific situation the obtained accuracy is the lowest one.

Up to this point, considering the achieved means, the largest one is obtained by FURIA, which also presented the largest result among all methods. The reverse occurs with the Baseline approach, which achieved accuracy means around 60%. A satisfactory performance was provided by the RIPPER method, around 70% of accuracy in general. Considering the C4.5 and RF the similarity of these approaches is also noticeable.

In a closer look to the obtained results, per hold-out, in the first analysis (a) the dominance of FURIA is observable, for 4/5 of the runs this approach achieved the largest mean. A similar situation also is noticeable, in the otter cases (b, c and d) since for 3 out 5 runs, FURIA achieved the largest mean. The Baseline is completely outperformed in two analysis, a and b, and for 4/5 and 3/5 for the analysis c and d respectively.

While RF achieve at least one satisfactory result in all analysis, C4.5 only have the largest accuracy in the one situation (d). It is interesting to mention that this last hold-out present a behavior in the last run where all the dataset performed equally.

4.1 Statistical Analysis

The analysis of the methods considering the accuracy is an interesting approach that demonstrated the superiority of the FURIA. However, the analysis of each hold-out and in a general way can not be enough to state any conclusion. In order to provide a complete study, a statistical group comparison is performed.

Precisely, the aligned Friedman rank test (Hodges and Lehmann, 1962) to compare the group of 5 different approaches, shows the achieved rankings per column in Table 5. Additionally, the values are sorted from the lowest to highest obtained ranking and is considered as control variable. We also compute the Holm's post-hoc test, to check whether the control approach is statistically better, showing the obtained APV with the obtained rank for each method. It there are statistical differences, considering a significance level of 10%, we underline it.

From the results obtained by the statistical analysis it can be concluded that FURIA is the best option to tackle with the presented issue. In fact, this method is considered as control variable in all situations.

Considering the obtained differences, we can observe that FURIA is statistically superior in relation to the Baseline in 80% of the study. Moreover, the performance is superior than C4.5 and RF in some situations as in *a* and *b*. In consideration of RIPPER, in all cases no differences were found.

4.2 Analyzing the Generated Models

As observable in the previous analysis, FURIA was the method that achieved the superior performance in the study. Thus, this subsection aims at analyzing the rules generated by this model.

Regarding the comprehension of the rules, its necessary to state that FURIA consider the usage of trapezoidal membership functions with a concept of soft boundaries. For example, as stated by the authors, a generic fuzzy rule R – $(A \in (-\infty, -\infty, 6, 9)|class_x)$ indicates that the rule is completely valid for A $\leq$ 6, invalid for A $>$ 9 and partially valid in-between.

In what follows we provide the rules generated by the model in the cases that it is considered as the largest global accuracy. That is, highlighted with $^{\uparrow}$ in Table 4. It is important to mention that the generated rules are the same for all cases. Also, consider CF as the confidence of the rule (Hühn and Hüllermeier, 2009).

- **[$Rule_1$] -** (TEMP_MAX in [32.5, 33.2, inf, inf]) and (ATMP_MEAN in [1014.46, 1014.5, inf, inf]) $\rightarrow$ LOSS_INTERVAL = B (CF = 0.63)
- **[$Rule_2$] -** (TEMP_MAX in [-inf, -inf, 32.5, 33.2]) $\rightarrow$ LOSS_INTERVAL = A (CF = 0.84)
- **[$Rule_3$] -** (PRECIPITATION in [-inf, -inf, 104.4, 105.6]) and (TOT_IN in [203680, 246120, inf, inf]) $\rightarrow$ LOSS_INTERVAL = A (CF = 0.9)

Table 4: Results achieved in test by the different approaches and validations.

Hold-out 75-25	FURIA	RIPPER	C4.5	RF	Baseline
Run 1	**0.746**	0.730	0.705	0.648	0.623
Run 2	**0.754**	0.697	0.730	0.689	0.598
Run 3	**0.770**	0.738	0.713	0.697	0.639
Run 4	**0.697**	0.689	0.664	0.631	0.582
Run 5	0.746	0.721	0.631	**0.762**$^{\uparrow}$	0.623
#Mean	**0.743**$^{\uparrow}$	0.715	0.689	0.685	0.613

(a)

Hold-out 80-20	FURIA	RIPPER	C4.5	RF	Baseline
Run 1	0.732	**0.742**	0.701	0.670	0.608
Run 2	**0.763**$^{\uparrow}$	0.753	0.742	0.711	0.577
Run 3	**0.763**	0.722	0.711	0.711	0.619$^{\downarrow}$
Run 4	0.660	**0.701**	0.639	0.639	0.577
Run 5	**0.742**	0.701	0.711	**0.742**	0.619
#Mean	**0.732**	0.724	0.701	0.695	0.600$^{\downarrow}$

(b)

Hold-out 85-15	FURIA	RIPPER	C4.5	RF	Baseline
Run 1	**0.767**$^{\uparrow}$	0.740	0.712	0.726	0.616
Run 2	**0.753**	0.671	0.726	0.671	0.575
Run 3	**0.795**$^{\uparrow}$	0.767	0.726	0.671	0.630
Run 4	0.644	**0.658**	0.616	0.548$^{\downarrow}$	0.616
Run 5	0.712	0.712	0.712	**0.740**	0.589$^{\downarrow}$
#Mean	**0.734**	0.710	0.699	0.671	0.605

(c)

Hold-out 90-10	FURIA	RIPPER	C4.5	RF	Baseline
Run 1	0.714	**0.735**	0.653	0.714	0.551$^{\downarrow}$
Run 2	**0.714**	0.673	0.694	0.673	0.531$^{\downarrow}$
Run 3	**0.735**	0.673	**0.735**	0.673	0.694
Run 4	0.633	0.673	0.612	0.612	**0.714**$^{\uparrow}$
Run 5	**0.735**	**0.735**	**0.735**	**0.735**	0.653
#Mean	**0.706**	0.698	0.686	0.682	0.629

(d)

Table 5: Statistical results with Align Friedman rank test and Holm post-hoc test.

Hold-out 75-25		
Method	Ranking	APV
FURIA	4	–
RIPPER	10.1	0.19
C4.5	13.7	0.08
RF	14.2	0.08
Baseline	23	0.00

(a)

Holdout 80-20		
Method	Ranking	APV
FURIA	5.6	–
RIPPER	7.7	0.65
C4.5	13.4	0.18
RF	15.3	0.11
Baseline	23	0.00

(b)

Hold-out 85-15		
Method	Ranking	APV
FURIA	1.4	–
RIPPER	2.6	0.33
C4.5	2.866667	0.23
RF	3.233333	0.08
Baseline	4.9	0.00

(c)

Holdout 90-10		
Method	Ranking	APV
FURIA	8.7	–
RIPPER	11.3	0.73
C4.5	12.9	0.73
RF	14.3	0.68
Baseline	17.8	0.2

(d)

- **[$Rule_4$] -** (RU_MIN in [-inf, -inf, 21, 24]) and (QTD_DAYS in [76, 78, inf, inf]) and (RU_MED in [-inf, -inf, 74.64, 74.66]) and (TEMP_MEAN in [-inf, -inf, 20.24, 20.25]) $\rightarrow$ LOSS_INTERVAL = C (CF = 0.62)

The rule generated by the algorithm in the $Rule_1$ states that when the maximum temperature is greater than 32.5°C (high temperatures) and the average atmospheric pressure is medium to high, the weight loss of the wood will be becoming considerable. Typically, maximum temperatures above 32.5°C occur in summer.

In the $Rule_2$, the algorithm says that when the maximum temperature of the period is up to 33.2ºC (high temperatures), the weight loss due to the moisture content will be considered small, with an 84% of confidence.

The rule generated in the $Rule_3$ states that when precipitation is low during the storage period, and the total input of wood logs in the pile is considered high, the weight loss due to moisture will be small. This rule is important, as it contradicts the relevant scientific literature, which establishes that rainfall (precipitation) is not a variable to be considered in the variation of moisture content. Based on the data considered in the development of the current research, the rainfall in the period is an important variable, with a confidence measure of 90%.

Finally, the $Rule_4$ establishes, with a confidence of 62%, that when the minimum relative humidity of the period is up to 24%; the number of days that the wood logs remain stored in the piles is high, from 76 days; and the average temperature of the period is up to 20.25ºC, the loss of moisture in the logs will be considerably high – when it should have a lower weight loss. What may have happened in the generation of this rule is a sample problem, as the generated result is different from the expected one.

5 CONCLUSION

Wood is a scarce resource. It serve as raw material for the industrialization of countless finished products, and even as fuel for factories, the use of this resource has grown in an increasing way around the world. However, after the tree is felled down, the wood gradually begins to lose its moisture content, causing cracks in the logs. Such cracks cause irreparable losses in production processes, as much wood that could be used in industry ends up being used very little – or even being discarded.

So, the objective of this work was to develop a method of predicting the moisture loss in wood while the logs are stored in piles – a step prior to industrialization, applying Machine Learning classification based methods to solve this problem.

Furthermore, this work compares and analyzes the results of the different applied algorithms: FURIA, Ripper, C4.5 and Random Forest. Of these, the classification method using the FURIA algorithm was superior to the others, including statistically superior in relation to the baseline.

From this paper, different future works can be consider. A tuning of the algorithms' hyperparameters as well as a regression approach.

ACKNOWLEDGEMENTS

This work was supported by the Brazilian research funding agencies CNPq (305805/2021-5) and FAPERGS (Programa de Apoio à Fixação de Jovens Doutores no Brasil - 23/2551-0000126-8).

REFERENCES

Allen, J. C. and Barnes, D. F. (1985). The causes of deforestation in developing countries. *Annals of the association of American Geographers*, 75(2):163–184.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Cohen, W. W. (1995). Fast effective rule induction. In *Twelfth International Conference on Machine Learning*, pages 115–123. Morgan Kaufmann.

Cordón, O., Del Jesus, M. J., and Herrera, F. (1999). A proposal on reasoning methods in fuzzy rule-based classification systems. *International Journal of Approximate Reasoning*, 20(1):21–45.

Estuqui Filho, C. A. (2006). A durabilidade da madeira na arquitetura sob a ação dos fatores naturais: estudo de casos em brasília. Master's thesis, Faculdade de Arquitetura e Urbanismo - Universidade de Brasília.

Freund, Y. and Mason, L. (1999). The alternating decision tree learning algorithm. In *icml*, volume 99, pages 124–133.

Harrington, P. (2012). *Machine learning in action*. Simon and Schuster.

Hodges, J. L. and Lehmann, E. L. (1962). Ranks methods for combination of independent experiments in analysis of variance. *Annals of Mathematical Statistics*, 33:482–497.

Hühn, J. and Hüllermeier, E. (2009). Furia: an algorithm for unordered fuzzy rule induction. *Data Mining and Knowledge Discovery*, 19(3).

Júnior, A. S. and Alves, J. E. (2019). Fatores intervenientes no armazenamento de laminados. *Gepros: Gestão da Produção, Operações e Sistemas*, 14(5):190.

Kengen, S. (2001). a política florestal brasileira: uma perspectiva histórica [brazilian forest policy: a historical perspective]. *1º Simpósio Ibero-americano de gestão e economia Florestal, Porto Seguro, Brazil. Instituto de Pesquisas e estudos Florestais.(2001). Retrieved May*, 4:2007.

Khorhidpoor, A., Nazari Shirkouhi, S., and Amin-Tahmasbi, H. (2023). Risk assessment of public-private partnership projects for water transmission and distribution using anfis method. *Sharif Journal of Industrial Engineering & Management*.

Lima, R., da Silva Cardoso, G., de Proença, G., and da Costa, W. G. (2017). Influência do tempo de armazenamento (tpc) da madeira no aceite de cavacos de eucalipto para a fabricação de polpa. In *Congresso Internacional de Celulose e Pape*, page 8. Associação Brasileira Técnica de Celulose e Papel – ABTCP.

Magalhães, V. M., Lucca, G., de Lima Bicho, A., and Borges, E. N. (2022). On the methods to predict moisture content on wood: A literature review. In *ICEIS (1)*, pages 521–528.

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, 9:381–386.

Quinlan, J. (1993). *C4.5: Programs for Machine Learning*. Morgan Kauffman.

Rezende, R. N., Lima, J. T., de Ramos, L. E., Faria, A. L. R., et al. (2010). Secagem ao ar livre de toras de eucalyptus grandis em lavras, mg. *Cerne*, 16:41–47.

Russell, S. J. (2010). *Artificial intelligence a modern approach*. Pearson Education, Inc.

Steinwart, I. and Christmann, A. (2008). *Support vector machines*. Springer Science & Business Media.

Tan, P.-N., Steinbach, M., and Kumar, V. (2005). *Introduction to Data Mining, (First Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Tavana, M., Nazari-Shirkouhi, S., Mashayekhi, A., and Mousakhani, S. (2022). An integrated data mining framework for organizational resilience assessment and quality management optimization in trauma centers. In *Operations Research Forum*, volume 3, page 17. Springer.

Tomczak, A., Grodziński, G., Jakubowski, M., Jelonek, T., and Grzywiński, W. (2018). Effects of short-term storage method on moisture loss and weight change in beech timber. *Croatian Journal of Forest Engineering: Journal for Theory and Application of Forestry Engineering*, 39(1):35–43.

Vieira, P. A., Buainain, A. M., and Figueiredo, E. V. C. (2019). O brasil alimentará a china ou a china engolirá o brasil? *Revista Tempo Do Mundo*, 2(1):51–82.

Yazdi, M. R. T., Mozaffari, M. M., Nazari-Shirkouhi, S., and Asadzadeh, S. M. (2018). Integrated fuzzy dea-anfis to measure the success effect of human resource spirituality. *Cybernetics and Systems*, 49(3):151–169.

Yegnanarayana, B. (2009). *Artificial neural networks*. PHI Learning Pvt. Ltd.