Article

Cut-to-Length Harvesting Prediction Tool: Machine Learning Model Based on Harvest and Weather Features

Rodrigo Oliveira Almeida [1,2,*], Richardson Barbosa Gomes da Silva [1] and Danilo Simões [1]

1 Department of Forest Science, Soils and Environment, School of Agriculture, São Paulo State University (UNESP), Botucatu 18610-034, Brazil; richardson.silva@unesp.br (R.B.G.d.S.); danilo.simoes@unesp.br (D.S.)
2 Federal Institute of Education, Science and Technology—Southeast of Minas Gerais (IFET), Muriaé 36884-036, Brazil
* Correspondence: rodrigooliveirafv@gmail.com

Abstract: Weather is a significant factor influencing forest health, productivity, and the carbon cycle. However, our understanding of these effects is limited for many regions and ecosystems. Assessing the impact of weather variability on harvester productivity from plantation forests may assist in forest planning through the use of data modeling. We investigated whether weather data combined with timber harvesting attributes could be used to create a high-performance model that could accurately predict harvester productivity in *Eucalyptus* plantations using machine learning. Furthermore, we aimed to provide an online application to assist forest managers in applying the model. For the modeling, we considered 15 weather and timber harvesting attributes. We considered productivity as the target attribute. We subjected the database to 24 common algorithms in default mode and compared them according to error metrics and accuracy. From the timber harvesting features combined with weather features, the Catboost model can predict the productivity of harvesters in a tuned mode, with a coefficient of determination of 0.70. The use of weather data combined with timber harvesting attributes in the model is an accurate approach for predicting harvester productivity in *Eucalyptus* plantations, allowing for the creation of an online, free application to assist forest managers.

Keywords: artificial intelligence; meteorological data mechanized; timber harvesting; *Eucalyptus* planted forests; forest operations

1. Introduction

The primary challenge for forest managers in forest-based industries is controlling productive inputs in a rational and efficient manner, thereby achieving high productivity. In addition to understanding the productive chain, plantation forest management requires a closer examination of potential factors that enhance the performance of harvesters. The increased data collection in harvesters and meteorological stations in plantation forests has attracted the attention of forestry companies, who have found machine learning techniques to be an efficient tool for planning operational activities.

Plantation forests are a type of forest that employs high-level management techniques, contains one or two species, and has trees with uniform spacing and age. As the demand for forest products continues to grow, plantation forests represent 7.0% of the total forested area worldwide, encompassing 294 million hectares. Brazil is well-known for its high productivity of *Eucalyptus* plantation forests, occupying 7.3 million hectares. This is largely due to the favorable climate for this species, as well as several plant breeding and plant management research initiatives [1–3].

In accordance with forest planning, the management of plantation forests encompasses the cultivation phase and the subsequent timber harvest. The timber harvest process is

defined as comprising a series of activities, including pruning, tracing, loading, transportation, and the provision of raw materials to the forest industry [4,5]. In the context of mechanized timber harvesting management, it is essential to select self-propelled forestry machines that are tailored to specific activities and functions. In Brazil, one of the primary timber-harvesting systems employed for *Eucalyptus* spp. is the cut-to-length system [6–8].

In the context of the cut-to-length system, the most prevalent self-propelled forestry machines are harvesters. These machines are capable of performing a diverse range of tasks, including felling, delimbing, topping, debarking, and bucking [9–13]. In this system, the productivity of the harvester is influenced by a multitude of factors, including the slope and soil type, the average volume, the density of the timber, the number of assortments, the technical characteristics of the self-propelled machine and cutting head, and the operator's experience level. In this context, the harvester provides several data points regarding the operation of harvest trees, which can be utilized to adjust forest management in order to achieve the optimal harvest yield [14,15].

In addition to the aforementioned factors, weather conditions can also affect harvester productivity, particularly for *Eucalyptus* spp. grown in plantations in more than 90 countries. Meteorological stations collect weather data, which are then used to determine synoptic atmospheric conditions. These data can be analyzed in relation to different research areas and crops, including *Eucalyptus* spp. [16–25].

The exponential growth in computational power has led to an exponential increase in the amount of data generated by mechanized timber harvesting. This has created a significant opportunity for the application of machine learning (ML) algorithms to these datasets, with the potential to enhance timber harvesting productivity and support more informed decision-making. However, the paucity of publications on the application of ML techniques to mechanized harvesting operations represents a significant obstacle to the advancement of this crucial field of research [26–29], particularly when it involves the integration of these data with climatic aspects [30].

Machine learning techniques, which may be classified as either supervised or unsupervised learning, are based on the extraction of patterns from a database in order to generate predictive models. These techniques are employed in a variety of studies within the field of forest research. In particular, regression analysis has been utilized in timber harvesting operations in order to predict productivity [31]. Ideally, models and experimental research should be closely integrated. A modeling framework can be employed to generate research questions and identify key sets of measurements needed. Furthermore, experimental data must be used critically to test model performance [32].

The objective of this study was to ascertain whether meteorological data from meteorological stations combined with timber harvesting attributes could be used to create a high-performance model that could accurately predict harvester productivity in *Eucalyptus* plantations using machine learning. Furthermore, the study aimed to provide an online application, accessible free of charge, to assist forest managers in applying the model.

2. Materials and Methods

2.1. Raw Data

We utilized structured data pertaining to mechanized timber harvesting operations in *Eucalyptus* plantations situated in two regions within the Brazilian state of Minas Gerais. The harvested timber was destined for paper and bleached *Eucalyptus* pulp production. The plantation area in question has a total area of 8609 hectares, with an average of 1028 trees ha^{-1}, a forest age of 7.5 years, an average individual tree volume of 0.20 m^3, and a slope terrain of 24°. The Köppen classification indicates that these forests are situated in a climate designated as Cwa, which is characterized by humid subtropical conditions with dry winters and hot summers [33]. The soils were classified as yellow Latosol, yellow red Argisol, and red Argisol, as described in the work of Santos et al. [34].

A cut-to-length system (CTL) was employed to harvest 1.77 million m^3 of timber over a 27-month period. This entailed a range of activities, including felling, delimbing,

topping, debarking, and bucking. The average log length was 6.4 m, and the work was carried out using 8-wheeled harvesters (model Ergo H8, Ponsse Plc, Vieremä, Finland) with a typical weight of 21,500 kg and an engine power of 205 kW. All machines utilized the harvester head (model H7 HD, Ponsse Plc, Vieremä, Finland) with a feed speed of 5 m s^{-1} and a maximum opening of 650 mm. To enhance the timber harvesting dataset, we employed weather data collected by the Brazilian National Institute of Meteorology [35] from automatic weather stations.

2.2. Data Processing

The raw dataset included the following features: timber assortment (CC—log with bark; EN—log-to-energy; and SC—log without bark), work shift (A—morning to afternoon; B—afternoon to night; and C—night to morning), working hours, operator experience (years), forest age (years), stand density (tree ha^{-1}), and average individual tree volume (m^3) obtained from a forest inventory. The mean wind speed (m s^{-1}), mean dew point temperature (Celsius degree), mean gust of wind (m s^{-1}), mean relative air humidity (%), mean atmospheric pressure (mbar), mean air temperature (Celsius degree), mean global radiation (kJ m^{-2}), and mean rainfall (mm) were also recorded by the meteorological stations. The target feature, productivity (m^3 h^{-1}), was also measured. The total volume harvested was determined using diameter and length sensors located on the harvester head. To prevent the under or overestimation of productivity, we validated the sensor accuracy using the methodology described by Santos et al. [36].

We applied several processes to the input dataset in order to produce meaningful data information. This process of data wrangling involves the removal of special characters, missing values, outliers, features without variation and incomplete data information. Furthermore, we employed exploratory data analysis (EDA) to ascertain the profile and distribution of the data, while recursive feature elimination (RFE) was utilized to retain the highly relevant features. We employed feature importance (FI) to remove features with low importance, as illustrated in the Supplementary Materials (Figures S1 and S2; Tables S1–S3), in order to facilitate the modeling process [37–39].

2.3. Modeling, Evaluation and Prediction

The input dataset generated was randomly divided into a training set and a test set according to an 8:2 ratio, resulting in 20,308 and 5078 instances, respectively. To ensure the data similarity between the two sets, we applied the EDA (Supplementary Materials—Table S4).

This study is based on the seven algorithm groups (Table 1) available for regression analysis via supervised learning. A total of 24 machine learning algorithms were applied with their default configurations in the Python programming language [40,41] with the objective of identifying the optimal models for each group (Supplementary Materials—Table S5).

Subsequently, the optimal default model was selected, and the optuna framework [42] was employed to adjust the hyperparameters, resulting in the tuned model (Supplementary Materials—Tables S6 and S7). The modeling process was conducted solely using data pertaining to timber harvesting features (THF) and timber harvesting features in conjunction with weather features (THFWF).

The metrics mean square error (MSE), mean absolute error (MAE), median absolute error (MedAE), maximum error (ME), and determination coefficient (R^2) were employed to assess the efficacy of the models based on the train and test datasets. In order to identify the optimal model, the Kruskal–Wallis test [43,44] was applied using the programming language R [45] (Supplementary Materials—Tables S8 and S9).

Two final models were applied to the test dataset to provide a comprehensive overview of the predictions (Supplementary Materials—Table S10), and we employed the SHapley Additive exPlanations (SHAP) method [46] to observe the effect of features over the models on prediction results. The general roadmap of the methodology is depicted in Figure 1.

Table 1. Main model groups that are possible to use in supervised machine learning for regression analysis. Groups 1 to 6 are provided by the Python scikit-learn module; Group 7 is provided by the Python gplearn module.

| Description	Model	Group
The target value is expected to be a linear combination of the features.	Linear Methods	1
In high-dimensional spaces, the use of a subset of training points in the decision function is an effective approach. The choice of kernel function is also a matter of discretion.	SVM	2
The prediction is based on the distance of the new point (Euclidean distance) to the samples analyzed, which are known as non-generalizing machine learning methods and non-parametric methods.	K nearest Neighbors	3
A non-parametric supervised learning method predicts the value of a target variable by learning simple decision rules inferred from the data features.	Decision Tree	4
The combination of the predictions of several estimators using a learning algorithm can enhance the generalizability and robustness of the resulting model over a single estimator.	Ensemble	5
The input and output layers may comprise one or more non-linear layers (hidden layers), which are capable of learning non-linear models and executing them in real time.	Artificial Neural Network	6
The identification of an underlying mathematical expression that best describes a relationship is a crucial initial step. This is followed by the construction of a population of naive random formulas, which represent the relationship between the known independent variables and their dependent variable targets. The objective is to predict new data.	Genetic Algorithm	7
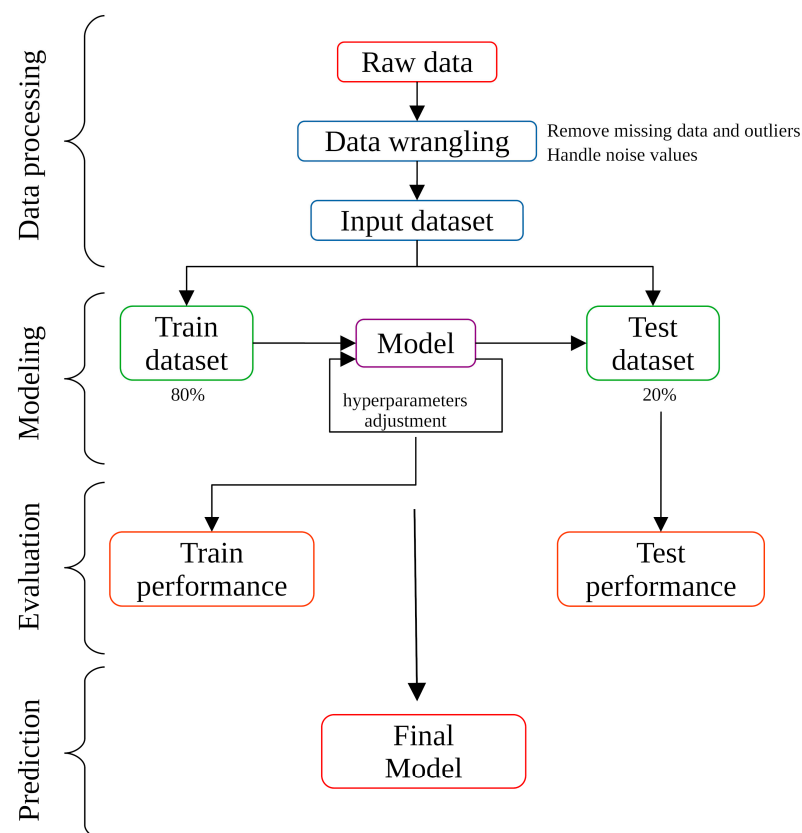
Figure 1. Flowchart of the methodology to generate the harvester productivity prediction model in timber-harvesting operations in *Eucalyptus* plantations.

3. Results

3.1. Recursive Feature Elimination and Feature Importance

The initial dataset comprised 35,223 instances and 43 features. Following the data wrangling process, the dataset pertaining to timber harvesting features (THF) now exhibits eight attributes, while the dataset incorporating timber harvesting features in conjunction with weather features (THFWF) displays sixteen attributes, both with 25,387 instances.

The initial features were ranked using the RFE method in the input dataset THF, with no feature being removed. In contrast, in the input dataset THFWF, the features mean air temperature, mean global radiation, and mean rainfall were removed. These results indicate that these features have low relevance in the dataset (Supplementary Materials—Table S2). Subsequently, the FI method demonstrated that the most crucial features were working hours and average individual tree volume, followed by forest age and operator experience, for both datasets (Figure 2 and Supplementary Materials—Table S3). Following the RFE method and FI, the THF dataset retained eight attributes, resulting in a total of 13 attributes in THFWF.
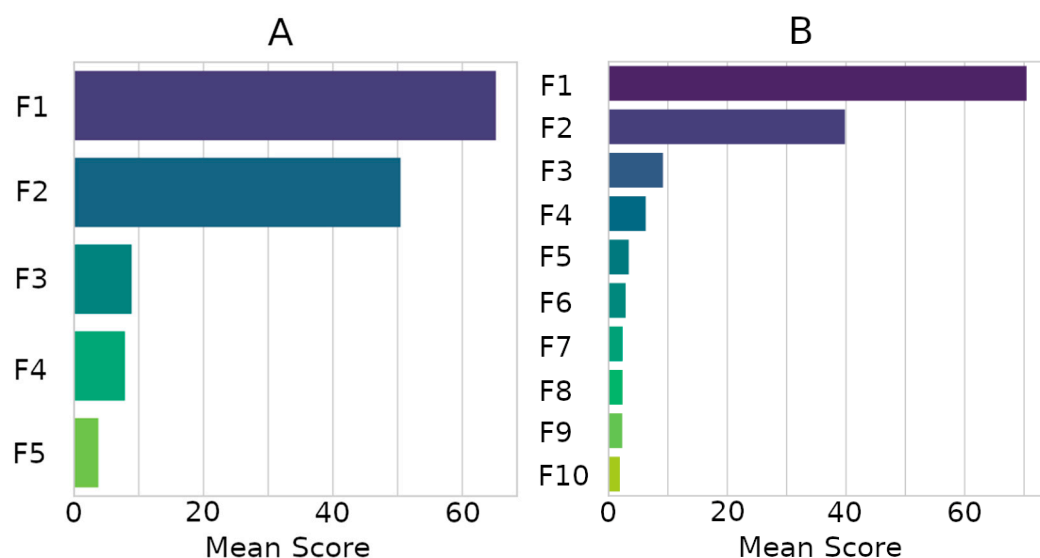
Figure 2. Feature importance analysis applied to database. (**A**) Only timber-harvesting features (THF): F1—working hours; F2—average individual tree volume (m^3); F3—forest age (years); F4—operator experience (years); F5—stand density (tree ha^{-1}); (**B**) timber harvesting features combined with weather features (THFWF): F1—working hours; F2—average individual tree volume (m^3); F3—forest age (years); F4—operator experience (years); F5—wind speed mean (m s^{-1}); F6—stand density (tree ha^{-1}); F7—mean atmospheric pressure (mbar); F8—mean relative air humidity (%); F9—mean dew point temperature (Celsius degree); F10—mean gust of wind (m s^{-1}).

3.2. Modeling in Default Mode

The ensemble, linear, and neural network methods (algorithm groups) yielded the most favorable outcomes in the training and test datasets across both input datasets, THF and THFWF (Supplementary Materials—Table S5). The CatBoost model demonstrated the most optimal performance in the test dataset, according to the Kruskal–Wallis test (Supplementary Materials—Table S8), with a higher value for R^2 and lower values for MSE and MAE (Tables 2–4).

Table 2. Mean value of the model performance by group of algorithms, analyzing the test dataset.

R^2	ME	MedAE	MAE	MSE	Group	
0.57	98.35	15.22	18.86	599.45	Linear Methods	
0.19	97.02	22.87	26.61	1119.18	SVM	
0.53	107.82	15.51	19.59	653	K Nearest Neighbors	
0.28	133.17	16.76	23.39	1001.68	Decision Tree	THF
0.63	101.75	13.34	17.16	513.29	Ensemble	
0.57	96.71	15.73	19.15	597.97	Artificial Neural Network	
0.54	112.42	13.73	18.69	638.22	Genetic algorithm	
0.58	103.75	14.13	18.29	577.82	Linear Methods	
0.24	97.49	24.87	26.74	1050.12	SVM	
0.35	109.85	20.18	23.97	904.21	K Nearest Neighbors	
0.29	128.07	16.91	23.19	982.87	Decision Tree	THFWF
0.65	102.48	13	16.73	485.92	Ensemble	
0.57	100.23	13.97	18.47	593.42	Artificial Neural Network	
0.55	119.67	13.04	18.43	628.27	Genetic algorithm	

Table 3. Best models for each group of models applied to test dataset analyzing THF data.

R^2	ME	MedAE	MAE	MSE	Model	Group
0.59	97.92	14.24	18.13	568.84	LassoLars	1
0.38	111.65	16.16	21.99	857.9	SVM Linear Kernel	2
0.53	107.82	15.51	19.59	653	K Nearest Neighbors	3
0.28	133.17	16.76	23.39	1001.68	CART	4
0.67	105.89	12.2	15.93	458.67	CatBoost	5
0.57	95.74	15.33	18.86	591.1	Multilayer Perceptron with 2 Layers	6
0.54	112.42	13.73	18.69	638.22	Symbolic Regressor	7

Table 4. Best models for each group of models applied to test dataset analyzing THFWF data.

R^2	ME	MedAE	MAE	MSE	Model	Group
0.59	104.59	13.88	18.09	567.81	Bayesian Ridge	1
0.48	112.73	20.24	22.25	720.55	SVM Linear Kernel	2
0.35	109.85	20.18	23.97	904.21	Nearest Neighbors	3
0.29	128.07	16.91	23.19	982.87	CART	4
0.69	105.38	11.77	15.4	425.99	CatBoost	5
0.57	99.33	14.04	18.5	593.15	Multilayer Perceptron with 2 Layers	6
0.55	119.67	13.04	18.43	628.27	Symbolic Regressor	7

3.3. Tuned Model

Following the hyperparameter adjustment process, the tuned model was created based on the best default model (CatBoost). According to the Kruskal–Wallis test (Supplementary Materials—Table S9), the tuned model (THF or THFWF data) demonstrated the highest performance on the test dataset. This was evidenced by the higher value of R^2 and the low values of MSE and MAE results (Tables 5 and 6).

Table 5. Performance of CatBoost model on test dataset THF data.

R^2	ME	MedAE	MAE	MSE	Mode	Final Model
0.67	105.89	12.2	15.93	458.67	Default	CatBoost
0.67	104.59	11.88	15.72	450.63	Tuned	

Table 6. Performance of CatBoost model on test dataset THFWF data.

R^2	ME	MedAE	MAE	MSE	Mode	Final Model
0.69	105.38	11.77	15.4	425.99	Default	CatBoost
0.7	109.09	11.57	15.18	414.5	Tuned	

3.4. SHAP Dependence Analysis

SHAP single dependency analysis was applied to the best-tuned model (with the use of weather data) to demonstrate the impact of the features in this model. The working hours and average individual tree volume were identified as the features with the greatest positive impact on the model output, while stand density, forest age, and timber assortment exhibited the greatest negative impact on the model output (Figure 3A,B).
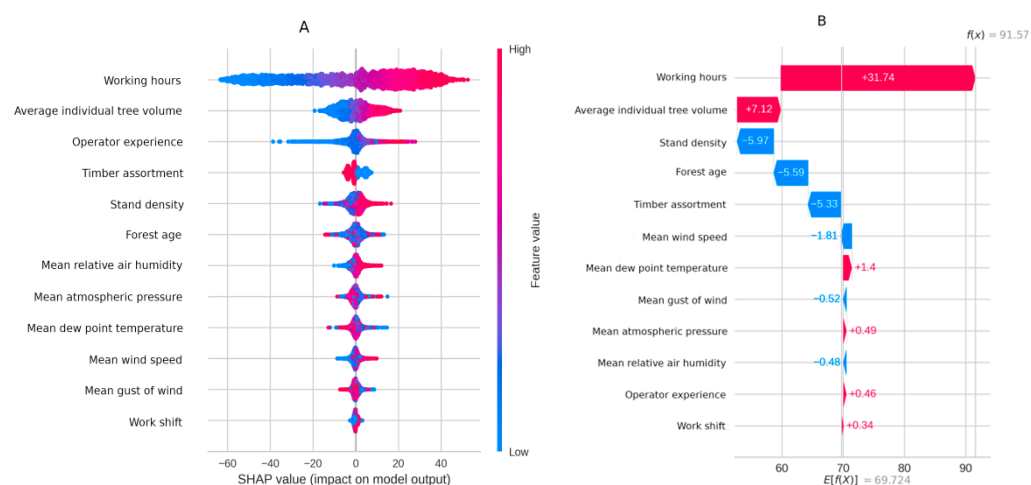
Figure 3. SHAP dependence analysis of the main features in harvester productivity data associated with weather data. (**A**) Impact of the feature values on model output; (**B**) impact of each feature (SHAP value) on the model output.

4. Discussion

4.1. Features in the Model

Irrespective of the dataset in question, namely THF or THFWF, the features pertaining to working hours, average individual tree volume, stand density, forest age, and timber assortment appear to be of particular relevance to this type of work. Indeed, these features exhibit a linear response to productivity. Some studies have indicated that tree volume and forest age play an important role in predictive models for harvester productivity [47,48].

However, our results indicate that the combination of harvest data with weather data produced the most optimal results, even if subtle, in the modeling process, as evidenced by Tables 2–4 and the Supplementary Materials (Tables S5 and S7). Despite exhibiting relatively low feature importance values, wind speed mean, mean atmospheric pressure, mean relative air humidity, mean dew point temperature, and mean gust of wind were able to enhance the predictive performance of the evaluated models. These features can impact the performance of harvester operators, such as impaired visibility due to sun glare or rain and exposure to different temperatures during entering/exiting the cabin or repair and maintenance work [49].

In terms of the utilization of features, the incorporation of weather data into the harvest data resulted in enhanced predictive models, which contributed to a reduction in the distortions and biases observed in models that solely utilized harvester data.

4.2. Predictive Models

An evaluation of the algorithms utilized in this study revealed that those belonging to the ensemble group exhibited the most favorable performance, regardless of whether the data in question were THF or THFWF. This observation was particularly evident in the case of the CatBoost algorithm, as evidenced by Tables 2–4. This algorithm has been successfully employed in a number of research endeavors, including those pertaining to power consumption forecasting and the estimation of building energy consumption. Its models have consistently demonstrated high performance [50,51]. Moreover, this type of model has demonstrated advantages in computational speed and reliable prediction when compared to other ensemble algorithms, such as XGBoost and LightGBM [52].

Following hyperparameter adjustments using the optuna framework to the default CatBoost model (generated using THF and THFWF data), it was found that the tuned CatBoost model exhibited the best performance. Consequently, this two-tuned model was selected as the final model. Nevertheless, as shown in Tables 5 and 6, the final model generated by the use of harvest data combined with weather data exhibited the lowest values for MSE and MAE, resulting in a higher value for R^2.

4.3. Impact of the Features on Predictive Model

The findings revealed that all features used in harvest data combined with weather data played a significant role in predicting productivity. The SHAP single dependency analysis method demonstrated the impact of these features on the final model.

When we consider the features working hours, average individual tree volume, mean dew point temperature, mean atmospheric pressure, operator experience, and work shift, it is evident that they exert a positive influence on the predictive model. On the other hand, the features stand density, forest age, timber assortment, mean wind speed, mean gust of wind, and mean relative air humidity exert a negative influence on the predictive model (Figure 3B). It can be concluded that the modifications in any features used in harvest data combined with weather data played a significant role, at different weights, in predicting productivity.

Operator experience in mechanized timber harvesting has been observed to increase over time, which has a positive impact on productivity [53]. Therefore, it is recommended that opportunities be provided to harvester operators in the form of adequate training, courses designed to transfer specific knowledge and standardization methods. These measures are likely to enhance the results of productivity. Furthermore, timber harvest data can be utilized to inform production control and forest management. These data can be employed to develop predictive models that assess harvester productivity, which can assist forest managers in making informed decisions [54]. Moreover, weather data collected by meteorological stations, whether by government institutions or private companies, can be utilized for a variety of purposes, including weather prediction and climate change studies [55].

4.4. Harvester Productivity Prediction Tool

Following the completion of the modeling process, the two models generated are available for download and use locally via the GitHub repository at https://github.com/AlmeidaRO/ML_Tools (accessed on 7 August 2024). Additionally, the same repository contains an online application based on Streamlit that can be accessed remotely [56], thus facilitating use by regular users. Streamlit is a Python-based tool for developing web applications for various purposes, primarily data science and machine learning. It enables users to rapidly and seamlessly construct web applications from Python scripts [57].

The online tool enables the user to select the model to be employed, with the models based on either timber harvesting feature (THF) or timber harvesting features combined with weather feature (THFWF) data. Consequently, in the absence of weather data, the forest manager must select the option "NO" in the weather data use option. The results

are displayed in real-time in response to the user's selection of an option in each feature (Figure 4).

Feature adjustments

Figure 4. Harvester Productivity Predictor, a user-friendly online tool to assist forest managers.

5. Conclusions

The application of machine learning techniques to the prediction of harvester productivity has been demonstrated to yield satisfactory results. The objective of our research was to integrate weather data with timber harvesting data in order to develop a more accurate predictive model for harvester productivity in *Eucalyptus* plantations. The integration of weather data into the model led to enhanced predictive capacity, indicating that the combination of these two data sources may be a valuable approach for enhancing the accuracy of harvester productivity predictions.

Moreover, the generated model has been employed to develop a harvester productivity prediction tool, an application (interactive dashboard) that enables the rapid visualization of harvester productivity in terms of specific situations and conditions. This tool is designed to assist forest managers in making informed decisions.

To direct further investigation, it is recommended that weather stations be installed in close proximity to regions with forest plantations. This will facilitate the creation of more precise models for predicting harvester productivity, thereby enabling the planning and implementation of forest harvesting operations. Furthermore, the utilization of this

localized meteorological data could enhance the accuracy of predictions generated by our model, which is accessible through the online application.

References

1. Sarre, A. *Global Forest Resources Assessment 2020: Main Report*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2020.
2. Stape, J.L.; Gonçalves, J.L.M.; Gonçalves, A.N. Relationships between Nursery Practices and Field Performance for *Eucalyptus* Plantations in Brazil. *New For.* **2001**, *22*, 19–41. [CrossRef]
3. Instituto Brasileiro de Geografia e Estatística. *Produção Da Extração Vegetal E Da Silvicultura*; IBGE: Rio de Janeiro, Brazil, 2022.
4. Alayet, C.; Lehoux, N.; Lebel, L.; Bouchard, M. Centralized Supply Chain Planning Model for Multiple Forest Companies. *INFOR Inf. Syst. Oper. Res.* **2016**, *54*, 171–191. [CrossRef]
5. Bligård, L.-O.; Häggström, C. CCPE—The Use of an Analytical Method to Evaluate Safety and Ergonomics in Maintenance of Forest Machinery. *Aust. For.* **2019**, *82*, 29–34. [CrossRef]
6. Camargo, D.A.; Munis, R.A.; Simões, D. Investigation of Exposure to Occupational Noise among Forestry Machine Operators: A Case Study in Brazil. *Forests* **2021**, *12*, 299. [CrossRef]
7. Palander, T.; Pasi, A.; Laurèn, A.; Ovaskainen, H. Comparison of Cut-to-Length Harvesting Methods in Tree Plantations in Brazil. *Forests* **2024**, *15*, 666. [CrossRef]
8. *ISO 6814:2016*; Machinery for forestry—Mobile and self-propelled machinery—Terms, definitions and classification. International Standardization for Organization: Rio de Janeiro, Brazil, 2016.

9.   Ackerman, S.A.; Talbot, B.; Astrup, R. The Effect of Tree and Harvester Size on Productivity and Harvester Investment Decisions. *Int. J. For. Eng.* **2022**, *33*, 22–32. [CrossRef]

10. Gagliardi, K.; Ackerman, S.A.; Ackerman, P.A. Multi-Product Forwarder-Based Timber Extraction: Time Consumption and Productivity Analysis of Two Forwarder Models Over Multiple Products and Extraction Distances. *Croat. J. For. Eng.* **2020**, *14*, 231–242. [CrossRef]

11. Karaszewski, Z.; Łacka, A.; Mederski, P.S.; Bembenek, M. Impact of Season and Harvester Engine RPM on Pine Wood Damage from Feed Roller Spikes. *Croat. J. For. Eng.* **2018**, *39*, 183–191.

12. Seng Hua, L.; Wei Chen, L.; Antov, P.; Kristak, L.; Tahir, P.M. Engineering Wood Products from *Eucalyptus* Spp. *Adv. Mater. Sci. Eng.* **2022**, *2022*, 8000780. [CrossRef]

13. Wessels, C.B.; Nocetti, M.; Brunetti, M.; Crafford, P.L.; Pröller, M.; Dugmore, M.K.; Pagel, C.; Lenner, R.; Naghizadeh, Z. Green-Glued Engineered Products from Fast Growing *Eucalyptus* Trees: A Review. *Eur. J. Wood Wood Prod.* **2020**, *78*, 933–940. [CrossRef]

14. Jankovský, M.; Merganič, J.; Allman, M.; Ferenčík, M.; Messingerová, V. The Cumulative Effects of Work-Related Factors Increase the Heart Rate of Cabin Field Machine Operators. *Int. J. Ind. Ergon.* **2018**, *65*, 173–178. [CrossRef]

15. Maldaner, L.F.; de Paula Corrêdo, L.; Canata, T.F.; Molin, J.P.; Felipe Maldaner, L.; de Paula Corrêdo, L.; Fernanda Canata, T.; Paulo Molin, J. Predicting the Sugarcane Yield in Real-Time by Harvester Engine Parameters and Machine Learning Approaches. *Comput. Electron. Agric.* **2021**, *181*, 105945. [CrossRef]

16. Booth, T.H. Eucalypt Plantations and Climate Change. *For. Ecol. Manag.* **2013**, *301*, 28–34. [CrossRef]

17. Câmara, A.P.; Vidaurre, G.B.; Oliveira, J.C.L.; Teodoro, P.E.; Almeida, M.N.F.; Toledo, J.V.; Júnior, A.F.D.; Amorim, G.A.; Pezzopane, J.E.M.; Campoe, O.C. Changes in Rainfall Patterns Enhance the Interrelationships between Climate and Wood Traits of *Eucalyptus*. *For. Ecol. Manag.* **2021**, *485*, 118959. [CrossRef]

18. Ding, X.; Zhao, Y.; Fan, Y.; Li, Y.; Ge, J. Machine Learning-Assisted Mapping of City-Scale Air Temperature: Using Sparse Meteorological Data for Urban Climate Modeling and Adaptation. *Build. Environ.* **2023**, *234*, 110211. [CrossRef]

19. de Freitas, E.C.S.; de Paiva, H.N.; Neves, J.C.L.; Marcatti, G.E.; Leite, H.G. Modeling of *Eucalyptus* Productivity with Artificial Neural Networks. *Ind. Crops Prod.* **2020**, *146*, 112149. [CrossRef]

20. Gao, W.; Shen, L.; Sun, S.; Peng, G.; Shen, Z.; Wang, Y.; Kandeal, A.A.W.; Luo, Z.; Kabeel, A.E.; Zhang, J.; et al. Forecasting Solar Still Performance from Conventional Weather Data Variation by Machine Learning Method. *Chin. Phys. B* **2023**, *32*, 048801. [CrossRef]

21. Martins, F.B.; Benassi, R.B.; Torres, R.R.; de Brito Neto, F.A. Impacts of 1.5 °C and 2 °C Global Warming on *Eucalyptus* Plantations in South America. *Sci. Total Environ.* **2022**, *825*, 153820. [CrossRef]

22. Queiroz, T.B.; Campoe, O.C.; Montes, C.R.; Alvares, C.A.; Cuartas, M.Z.; Guerrini, I.A. Temperature Thresholds for *Eucalyptus* Genotypes Growth across Tropical and Subtropical Ranges in South America. *For. Ecol. Manag.* **2020**, *472*, 118248. [CrossRef]

23. Rocha, S.M.G.; Vidaurre, G.B.; Pezzopane, J.E.M.; Almeida, M.N.F.; Carneiro, R.L.; Campoe, O.C.; Scolforo, H.F.; Alvares, C.A.; Neves, J.C.L.; Xavier, A.C.; et al. Influence of Climatic Variations on Production, Biomass and Density of Wood in *Eucalyptus* Clones of Different Species. *For. Ecol. Manag.* **2020**, *473*, 118290. [CrossRef]

24. Sondermann, M.; Chou, S.C.; Lyra, A.; Latinovic, D.; Siqueira, G.C.; Junior, W.C.; Giornes, E.; Leite, F.P. Climate Change Projections and Impacts on the *Eucalyptus* Plantation around the Doce River Basin, in Minas Gerais, Brazil. *Clim. Serv.* **2022**, *28*, 100327. [CrossRef]

25. Yang, L.; Kong, J.; Gao, Y.; Chen, Z.; Lin, Y.; Zeng, S.; Su, Y.; Li, J.; He, Q.; Qiu, Q. A Simulated Drier Climate Reduces Growth and Alters Functional Traits of *Eucalyptus* Trees: A Three-Year Experiment in South China. *For. Ecol. Manag.* **2023**, *549*, 121435. [CrossRef]

26. Lopes, I.L.E.; Araújo, L.A.; Miranda, E.N.; Bastos, T.A.; Gomide, L.R.; Castro, G.P. A Comparative Approach of Methods to Estimate Machine Productivity in Wood Cutting. *Int. J. For. Eng.* **2022**, *33*, 43–55. [CrossRef]

27. Melander, L.; Ritala, R. Separating the Impact of Work Environment and Machine Operation on Harvester Performance. *Eur. J. For. Res.* **2020**, *139*, 1029–1043. [CrossRef]

28. Munis, R.A.; Almeida, R.O.; Camargo, D.A.; da Silva, R.B.G.; Wojciechowski, J.; Simões, D. Machine Learning Methods to Estimate Productivity of Harvesters: Mechanized Timber Harvesting in Brazil. *Forests* **2022**, *13*, 1068. [CrossRef]

29. Svoikin, F.; Zhuk, K.; Svoikin, V.; Ugryumov, S.; Bacherikov, I.; Iniesta, D.V.; Ryapukhin, A. Classification of Tree Species in the Process of Timber-Harvesting Operations Using Machine-Learning Methods. *Inventions* **2023**, *8*, 57. [CrossRef]

30. Elli, E.F.; Sentelhas, P.C.; Bender, F.D. Impacts and Uncertainties of Climate Change Projections on *Eucalyptus* Plantations Productivity across Brazil. *For. Ecol. Manag.* **2020**, *474*, 118365. [CrossRef]

31. Munis, R.A.; Almeida, R.O.; Camargo, D.A.; da Silva, R.B.G.; Wojciechowski, J.; Simões, D. Tactical Forwarder Planning: A Data-Driven Approach for Timber Forwarding. *Forests* **2023**, *14*, 1782. [CrossRef]

32. Medlyn, B.E.; Duursma, R.A.; Zeppel, M.J.B. Forest Productivity under Climate Change: A Checklist for Evaluating Model Studies. *WIREs Clim. Chang.* **2011**, *2*, 332–355. [CrossRef]

33. Alvares, C.A.; Stape, J.L.; Sentelhas, P.C.; de Moraes Gonçalves, J.L.; Sparovek, G. Köppen's Climate Classification Map for Brazil. *Meteorol. Z.* **2013**, *22*, 711–728. [CrossRef]

34. dos Santos, H.G.; Jacomine, P.K.T.; dos Anjos, L.H.C.; de Oliveira, V.A.; Lumbreras, J.F.; Coelho, M.R.; de Almeida, J.A.; de Araujo Filho, J.C.; de Oliveira, J.B.; Cunha, T.J.F. *Sistema Brasileiro de Classificação de Solos*, 5th ed.; Embrapa: Brasília, Brazil, 2018; ISBN 978-85-7035-800-4.

35. Instituto Nacional de Meteorologia. Available online: https://bdmep.inmet.gov.br/ (accessed on 20 January 2023).

36. Santos, D.W.F.D.N.; Magalhães Valente, D.S.; Fernandes, H.C.; Souza, A.P.D.; Cecon, P.R. Modeling Technical, Economic and Environmental Parameters of a Forwarder in a *Eucalyptus* Forest. *Int. J. For. Eng.* **2020**, *31*, 197–204. [CrossRef]

37. George, A. Anomaly Detection Based on Machine Learning Dimensionality Reduction Using PCA and Classification Using SVM. *Int. J. Comput. Appl.* **2012**, *47*, 5–8. [CrossRef]

38. Guyon, I.; Elisseeff, A. An Introduction to Variable and Feature Selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.

39. Reddy, G.T.; Reddy, M.P.K.; Lakshmanna, K.; Kaluri, R.; Rajput, D.S.; Srivastava, G.; Baker, T. Analysis of Dimensionality Reduction Techniques on Big Data. *IEEE Access* **2020**, *8*, 54776–54788. [CrossRef]

40. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python Fabian. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

41. Stephens, T. Gplearn: Genetic Programming in Python, with a Scikit-Learn Inspired API. Available online: https://github.com/trevorstephens/gplearn (accessed on 20 January 2023).

42. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-Generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019. [CrossRef]

43. Kruskal, W.H.; Wallis, W.A. Use of Ranks in One-Criterion Variance Analysis. *J. Am. Stat. Assoc.* **1952**, *47*, 583. [CrossRef]

44. de Mendiburu, F.; Yaseen, M. Agricolae: Statistical Procedures for Agricultural Research. Available online: https://myaseen208.github.io/agricolae/ (accessed on 1 January 2023).

45. R Core Team. R: A Language and Environment for Statistical Computing. Available online: https://www.r-project.org/ (accessed on 1 January 2023).

46. Lundberg, S.M.; Lee, S.-I.I. A Unified Approach to Interpreting Model Predictions. In *Proceedings of the Advances in Neural Information Processing Systems*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: New York, NY, USA, 2017; Volume 30.

47. Liski, E.; Jounela, P.; Korpunen, H.; Sosa, A.; Lindroos, O.; Jylhä, P. Modeling the Productivity of Mechanized CTL Harvesting with Statistical Machine Learning Methods. *Int. J. For. Eng.* **2020**, *31*, 253–262. [CrossRef]

48. Louis, L.T.; Kizha, A.R.; Daigneault, A.; Han, H.-S.; Weiskittel, A. Factors Affecting Operational Cost and Productivity of Ground-Based Timber Harvesting Machines: A Meta-Analysis. *Curr. For. Rep.* **2022**, *8*, 38–54. [CrossRef]

49. Häggström, C.; Lindroos, O. Human, Technology, Organization and Environment—A Human Factors Perspective on Performance in Forest Harvesting. *Int. J. For. Eng.* **2016**, *27*, 67–78. [CrossRef]

50. Ke, J.; Qin, Y.; Wang, B. Optimizing and Controlling Building Electric Energy Using Cat Boost Under the Energy Internet of Things. In Proceedings of the 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2), Wuhan, China, 30 October–1 November 2020; pp. 553–558.

51. Pan, Y.; Zhang, L. Data-Driven Estimation of Building Energy Consumption with Multi-Source Heterogeneous Data. *Appl. Energy* **2020**, *268*, 114965. [CrossRef]

52. Qian, L.; Chen, Z.; Huang, Y.; Stanford, R.J. Employing Categorical Boosting (CatBoost) and Meta-Heuristic Algorithms for Predicting the Urban Gas Consumption. *Urban Clim.* **2023**, *51*, 101647. [CrossRef]

53. Jain, A. How Knowledge Loss and Network-Structure Jointly Determine R&D Productivity in the Biotechnology Industry. *Technovation* **2023**, *119*, 102607. [CrossRef]

54. Rosińska, M.; Bembenek, M.; Picchio, R.; Karazzewski, Z.; Đuka, A.; Mederski, P.S.; Karaszewski, Z.; Đuka, A.; Mederski, P.S. Determining Harvester Productivity Curves of Thinning Operations in Birch Stands of Central Europe. *Croat. J. For. Eng.* **2022**, *43*, 1–12. [CrossRef]

55. Cock, J.; Jiménez, D.; Dorado, H.; Oberthür, T. Operations Research and Machine Learning to Manage Risk and Optimize Production Practices in Agriculture: Good and Bad Experience. *Curr. Opin. Environ. Sustain.* **2023**, *62*, 101278. [CrossRef]

56. Streamlit—A Faster Way to Build and Share Data Apps. Available online: https://streamlit.io/ (accessed on 1 November 2023).

57. Lee, C.; Lin, J.; Prokop, A.; Gopalakrishnan, V.; Hanna, R.N.; Papa, E.; Freeman, A.; Patel, S.; Yu, W.; Huhn, M.; et al. StarGazer: A Hybrid Intelligence Platform for Drug Target Prioritization and Digital Drug Repositioning Using Streamlit. *Front. Genet.* **2022**, *13*, 868015. [CrossRef]