forests

MDPI

Article

Tactical Forwarder Planning: A Data-Driven Approach for Timber Forwarding

Rafaele Almeida Munis [1], Rodrigo Oliveira Almeida [1], Diego Aparecido Camargo [1], Richardson Barbosa Gomes da Silva [1], Jaime Wojciechowski [2] and Danilo Simões [1,*]

[1] Department of Forest Science, Soils and Environment, School of Agriculture, São Paulo State University (UNESP), Botucatu 18610-034, Brazil; rafaele.munis@gmail.com (R.A.M.); rodrigooliveiraufv@gmail.com (R.O.A.); diegocamargoflorestal@gmail.com (D.A.C.); richardson.silva@unesp.br (R.B.G.d.S.)

[2] Informatics Department, Federal University of Paraná, Curitiba 81520-260, Brazil; jaimewo@ufpr.br

* Correspondence: danilo.simoes@unesp.br

Abstract: Tactical planning in timber harvesting involves aspects related to forest macro-planning and, particularly, the allocation of resources and sequencing of activities, all of which affect the allocation of timber in forest yards and roads and the productivity of forest machines. Data-driven approaches encourage the use of information obtained from data to enhance decision-making efficiency and support the development of short-term strategies. Therefore, our investigation was intended to determine whether a data-driven approach can generate sufficient input for modeling forwarder productivity in timber forwarding in *Pinus* and *Eucalyptus* planted forests, to support tactical planning. We utilized 3812 instances of raw data that were generated over a 36-month period. The data were collected from 23 loggers who operated in *Pinus* and *Eucalyptus* planted forests. We applied 22 regression algorithms that applied a supervised learning method from an experimental machine learning approach to the data instances. We evaluated the fitted models using three performance metrics. Out of the tested algorithms, the default mode of light gradient boosting produced a root mean squared error of 14.80 m^3 h^{-1}, a mean absolute error of 2.70, and a coefficient of determination of 0.77. Therefore, data-driven methods adequately support forwarder productivity modeling in timber forwarding in planted forests and help forest managers with tactical planning.

Keywords: predictive analytics; forest operations; machine learning; forest productivity; intelligent forest management; cut-to-length harvesting

Citation: 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. https://doi.org/10.3390/f14091782

Received: 3 July 2023
Revised: 29 August 2023
Accepted: 29 August 2023
Published: 1 September 2023

1. Introduction

Data manipulation techniques, such as machine learning, can enable integrated tactical planning of mechanized harvesting, timber transfer to storage yards, and forest roads. Integrating these processes promotes rational resource management, efficiently allocating raw materials to meet the needs of forest-based industries. Forest managers should not only comprehend the production chain of planted forests but also be mindful of the increments in the performance standards of the man–machine–machine system. Evaluating the interactions between the system and the external environment is crucial to identifying and enhancing the operator's functions during the process of integrating a forest machine.

Economically, revenue should offset the opportunity costs incurred and cover the expenses of mechanized systems chosen to ensure a project's viability [1]. The financial costs involved in implementing forest management strategies in timber harvesting systems can be evaluated by constructing predictive models to determine their feasibility. For instance, when adopting a harvesting system, these models offer an advanced classification of the conditions it will face.

The cut-to-length harvesting system has become the most popular method in the paper industry. This system typically involves harvesters and forwarders that process the trees in the harvest block [2,3]. Mapping the productivity of harvesting forest machines, such as the forwarder, in cut-to-length systems can be a valuable source of data for smart forest management. These data accounts for all events and interventions during timber harvesting and can be used to inform forest management strategies [4].

The productivity mapping of forwarders provides forest managers with quantitative data that can be conveniently used for tactical planning. The use of ground machines to harvest timber is common practice worldwide in forestry operations. Therefore, the performance of a harvesting team depends on discretionary factors controlled by the unit, such as machine and operator selection, which affect its ability to convert inputs into outputs [5–7].

Forest managers can anticipate challenges, allocate resources, and establish objectives using the three levels of decision making in a traditional pyramid scheme: strategic, tactical, and operational decision making. However, at a tactical level, forest managers evaluate the mid-term decisions needed to make sustainable decisions, ensure effective raw material supply, and execute operational plans [8–11].

Forest managers can strategize harvesting to meet specific unit objectives within a given time scale by using data-driven approaches—such as collecting data in response to a specific demand—when sizing the fleet of forwarders to be allocated. Using data-driven approaches can reduce or eliminate subjectivity in parameter assignment, leading to reliable and efficient forecasts [12–14].

Machine learning techniques use a data-driven approach to generate reliable information and speed up the process of obtaining performance metrics for forwarder activities in planted forests. Unfortunately, traditional approaches for obtaining information on the productivity of forest harvesting machines can be subjective; this is mainly because they are not generalizable. When it comes to efficiency and accuracy, data-driven procedures have an advantage over traditional, experience-based approaches [15–17].

To achieve high accuracy during analysis, machine learning techniques are commonly employed to create analysis methodologies that identify patterns or trends. These techniques are utilized to perform critical functions in data mining, text mining, predictive analytics, and decision-making systems [18–20]. One type of machine learning is the supervised approach. When using the supervised approach, algorithms are trained using labeled data to make predictions or decisions. The supervised approach has several advantages in data analysis scenarios, as it allows for highly accurate predictions, adaptation to complex problems, and decision optimization. In supervised learning, the model adjusts its parameters iteratively to match the output of the predictions based on the input, ensuring accuracy [21,22].

The literature has discussed the use of machine learning for predicting the productivity of forest machines used in mechanized timber harvesting. Eriksson and Lindroos [23] trained machine learning models to estimate the productivity of harvesters and forwarders. Estimation accuracy was higher for harvesters. Guerra et al. [24] estimated the productivity of harvesters that operate in *Pinus* and *Eucalyptus* forests. Ghaffariyan et al. [25] predicted the productivity of harvesters and forwarders using multiple regression models. The focus was on the explanatory power of the predictor variables.

However, forecasting and monitoring forwarder productivity can provide forest managers with insights into demand forecasting, routing, behavioral analysis, facility planning, and, most importantly, truck traffic forecasting [26]. A gap in the literature was detected regarding the utilization of diverse machine learning techniques and empirical data pertaining to forest harvesting machines, planted forests, and working shifts in the estimation of forwarder productivity. As forest managers plan to harvest planted forests, a predictive model of forwarder productivity is expected to support the tactical planning of operations, including the dimensioning of the schedule of activities, shifts of operations, and the number of forwarders required. We investigated whether a data-driven approach, specifically machine learning, can provide sufficient information for predicting productivity in timber forwarding in *Pinus* and *Eucalyptus* planted forests. Our investigation aimed to determine whether a data-driven approach can generate sufficient input for modeling forwarder productivity in timber forwarding in *Pinus* and *Eucalyptus* planted forests to support tactical planning.

2. Materials and Methods

2.1. Data Origin

The study collected empirical data from the controller area network (CAN) of timber logging activity carried out by forwarders after timber harvesting had been conducted by harvesters in Brazilian *Pinus* and *Eucalyptus* planted forests. The total areas of these forests were 325 ha and 290 ha for *Pinus* and *Eucalyptus*, respectively. The region's slopes range from 7.32% to 35.06%, and the intervals have been classified by Speight and Isbell [27] as gentle (3% to 10%), moderate (10% to 32%), and steep (32% to 56%) terrain.

The forests planted with *Pinus* had a spacing of 3.3 m $\times$ 1.8 m, an average age of 22 years, and a total harvested volume of 146,680 m^3. Forests planted with *Eucalyptus* also had a spacing of 3.3 m $\times$ 1.8 m, but the average age was 14 years, and the total harvested volume was 235,720 m^3. The average distance for timber transportation from the *Pinus* forest was 273 m, and for the *Eucalyptus* forest, this was 285 m. After harvesting, the timber was placed at the edge of the stands near forest roads for later loading and transportation to the forest-based industry for paper production.

We analyzed the operational data of 21 Elephant King forwarders (model manufactured by Ponsse Plc, Vieremä, Finland) under normal working conditions. The forwarders were rated at 205 kW, with a load capacity of 20,000 kg, and were equipped with a 0.36 m^2 tire wheel systems and 8$\times$8 traction. Additionally, we analyzed two Ponsse Buffalo forwarders (model manufactured by Ponsse Plc, Vieremä, Finland) with a rating of 210 kW and a capacity of 15,000 kg. The forwarders were equipped with a 0.36 m^2 grab, tire wheel systems, and 8 $\times$ 8 traction.

2.2. Dependent and Target Variable Dataset

We used 3812 instances of raw data generated over a 36-month period. Since yield prediction can be influenced by different forwarder attributes [28], the evaluated dataset consisted of information related to forwarders, planted forests, and working shifts. The analysis employed six variables, of which four were continuous (Table 1) and two were categorical. The target variable was the productive machine hours (m^3 h^{-1}). It should be noted that the forwarder hour meter was included in the model adjustments but was ultimately removed due to it increasing the complexity of the adjustment without contributing to the minimization of the error.

Table 1. Data profiling of the dataset used to fit the forwarder productivity prediction models.

Features	Mean	Standard Deviation	Minimum Value	Median	Maximum Value
Terrain slope (%)	12.85	5.81	4.49	11.84	37.46
Plot area (m^2)	19.48	19.20	0.11	12.42	76.71
Total distance (m)	5387.78	2125.68	-	5380.50	17,269.00
Average distance (m)	281.41	170.68	-	231.79	1489.50

We evaluated the total and average timber transport distance (expressed in meters) among the dependent variables related to the forwarders. The attributes related to the planted forests that we evaluated included the terrain slope (%), plot area (m^2), and species. To meet the industrial demand for timber in Brazil, two daily working shifts are usually necessary. In our analysis, we also considered working shifts, which we transformed into a dummy variable with a value of zero or one to indicate the absence or presence of the attribute. This allowed us to classify the data into mutually exclusive categories, each consisting of 10 h per day.

Forwarder productivity was estimated as the product of the maximum timber capacity carried by the forwarder and the reported number of loads forwarded, divided by the effective time of the activity. Thus, productivity (Equation (1)) is calculated by dividing the total timber forwarded (υ) in m^3 by the corresponding sum of productive machine hours (h).

$$\text{productivity} = \frac{\upsilon}{h} \quad (1)$$

After obtaining the raw data, we utilized R to clean and transform the dataset, a process known as data wrangling [29]. The majority of variables were utilized in the model in their original state, with the exception of the variable 'working shifts', which was converted to dummy variables in order to evaluate their impact. Moreover, outliers were eliminated through the application of the interquartile range method, which was conceptually outlined using the Tukey range, and the missing values of numerical variables were also addressed. Additional information is available in Supplementary Materials (Figure S1).

2.3. Forwarder Productivity Modeling

In the first modeling step, the supervised algorithms are applied to the training data set in default mode. From the data instances, we assigned 80% (n = 3049) as a training set and 20% (n = 763) as the testing set.

We would like to note that the supervised machine learning technique comprises a series of ordered steps necessary to accomplish the prediction model. These steps include: determining the type of study—classification, regression, or anomaly detection; performing data cleaning and processing steps and transforming the data to ensure it is prepared for analysis; dividing the data into training, validation, and testing sets; identifying the variables in the dataset that will serve as inputs to the model; selecting one or more machine learning algorithms suitable for the problem; adjusting the model hyperparameters using the validation set (if applicable) and evaluating performance on unseen data; utilizing appropriate evaluation metrics to measure the model's performance on the test data.

The weights assigned to the regression models are determined based on their accuracy. The final prediction is obtained iteratively through a workflow that consolidates the predictions of regression models applied to the test dataset [30,31]. In our supervised learning framework, labeled data were randomly arranged in the training set to update the parameters. As a result, the labels were displayed during training and masked during testing.

A forwarder productivity prediction model was built using the scikit-learn library [32] in Python. Machine learning algorithms were applied to the data through supervised learning [33,34]. The adjustment of hyperparameters was evaluated to optimize the folds, iterations, and estimator of the model with different learning methods such as default mode, tuned mode, voting mode, and stacking mode [35–38].

By default, algorithms have limited impact because they do not optimize their hyperparameter configurations [39]. To overcome this problem, the model's parameters are adjusted—for example, with Tuned—to optimize its performance. This process involves finding the optimal combination, which leads to better performance.

To develop models capable of predicting forwarder productivity and optimizing their output, hyperparameter tuning is necessary. Hyperparameter tuning is the automatic optimization of a machine learning model's parameters [36].

However, several advanced techniques involving ensemble learning are evaluated, including voting and stacking. The voting weighting system enables two or three machine learning methods to complement each other [37,40]. A stacked model enables the combination of various machine learning algorithms under the supervision of the metamodel, ultimately achieving the highest attainable accuracy [41].

Machine learning is an experimental process. In other words, when we build a machine learning model, we allow computer programs to learn from the provided data and make adjustments to the hyperparameters based on the same data [42]. We evaluated several algorithms, hyperparameters, and learning methods to identify the most effective solution.

We tested the following 22 machine learning algorithms with various methodological bases to determine the best fit for estimating forwarder productivity, as suggested by Souza et al. [43]. This study utilized a total of 22 algorithms, chosen for their ease of implementation and ability to execute parallel processing (multithreading) using the Python libraries scikit-learn, CatBoost, XGBoost, and SymbolicRegressor [32,44,45]. These algorithms belong to the category of linear methods (Automatic Relevance Determination, Bayesian Ridge, ElasticNet, Huber, Lasso, LassoLars, Linear Regression, Orthogonal Matching Pursuit, Ridge), decision tree (CART), nearest neighbors (Nearest Neighbors), support vector machine (Support Vector Machine Linear Kernel, SVM rbf Kernel), neural artificial network (Multilayer Perceptron) and ensemble methods (AdaBoost, Extremely Randomized Trees, CatBoost, Extreme Gradient Boosting, Gradient Boosting, Light Gradient Boosting, Random Forest) and genetic algorithm (Symbolic Regressor).

2.4. Choosing the Best Model

To evaluate the performance of the fitted models, we used three commonly recommended quantitative methods [46,47]. We used the coefficient of determination (R^2) of the predicted (P_i) and observed (S_i) values to express the variation in the response variable over the variation in the predictor variables (Equation (2)) [48]:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(P_i - S_i\right)^2}{\sum_{i-1}^{n}\left(P_i - \bar{S}\right)^2} \tag{2}$$

We used the root mean square error (RMSE) to measure the Euclidean distance between the actual values and the predicted values (Equation (3)) [49]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - S_i)^2}{n}} \tag{3}$$

We used the mean absolute error (MAE) to measure the error between paired observations (Equation (4)) [50]:

$$MAE = \frac{\sum_{i=1}^{n}|P_i - S_i|}{n} \tag{4}$$

3. Results

The performance of each model was evaluated by comparing the predicted values to the actual yield values. In default mode, the models produced an average variance of 0.56 ± 0.22, as shown in Figure 1. It is noteworthy that the R^2 of Light Gradient Boosting exceeded the general average R^2 by more than 21%.

Four boosting algorithms (Light Gradient boosting), two decision tree algorithms (Random Forest), and one clustering algorithm (Nearest Neighbors) generated the fewest prediction errors in the default mode of the test dataset (Table 2).

Table 2. Prediction errors in terms of the root mean square error (RMSE) and mean absolute error (MAE) generated by the models in the default mode.

Model	RMSE (Mean ± Std) [m^3 h^{-1}]	MAE (Mean ± Std)
Light Gradient Boosting	14.80 ± 10.55	2.70 ± 0.86
Gradient Boosting	16.11 ± 10.55	2.81 ± 0.87
CatBoost	16.89 ± 11.64	2.83 ± 0.87
Extreme Gradient Boosting	16.97 ± 11.99	2.89 ± 0.87
Random Forest	18.37 ± 13.10	3.00 ± 0.95
Extremely Randomized Trees	19.23 ± 13.39	3.01 ± 1.04

Table 2. *Cont.*

Model	RMSE (Mean ± Std) [m^3 h^{-1}]	MAE (Mean ± Std)
Nearest Neighbors	20.34 ± 12.31	3.33 ± 1.01
AdaBoost	23.11 ± 8.95	3.83 ± 0.79
Multilayer Perceptron with One Layer	25.82 ± 14.69	3.64 ± 0.99
Lasso Regression	27.21 ± 11.39	4.16 ± 0.85
Symbolic Regressor	27.49 ± 22.90	3.28 ± 1.25
ElasticNet	27.57 ± 12.72	4.21 ± 0.89
Orthogonal Matching Pursuit	27.74 ± 13.20	4.20 ± 0.91
Bayesian Ridge	28.20 ± 12.36	4.28 ± 0.86
Ridge Regression	28.27 ± 12.63	4.27 ± 0.96
LassoLars	28.79 ± 10.80	4.33 ± 0.82
Linear Regression	28.82 ± 12.88	4.29 ± 0.90
Huber Regression	31.24 ± 12.50	4.27 ± 0.92
CART	34.43 ± 21.89	4.02 ± 1.27
Automatic Relevance Determination	37.70 ± 11.50	5.02 ± 0.80
SVM rbf Kernel	38.02 ± 13.38	4.97 ± 0.87
Multilayer perceptron with Two Layers	38.86 ± 18.38	4.77 ± 1.11
SVM Linear Kernel	68.01 ± 20.29	7.15 ± 1.20

We controlled the strength of the penalty parameters of the Ridge Regression and Lasso Regression algorithms, which showed the best performance in the default mode, through a tuned process (Table 3).

Table 3. Performance of the best models in the default mode after tuning the process coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE).

Model	R^2 (Mean ± Std)	RMSE (Mean ± Std) [m^3 h^{-1}]	MAE (Mean ± Std)
Light Gradient Boosting	0.74 ± 0.18	16.63 ± 12.16	2.87 ± 0.83
Nearest Neighbors	0.72 ± 0.18	17.68 ± 11.39	3.13 ± 0.85
Random Forest	0.69 ± 0.17	19.67 ± 11.34	3.32 ± 0.94

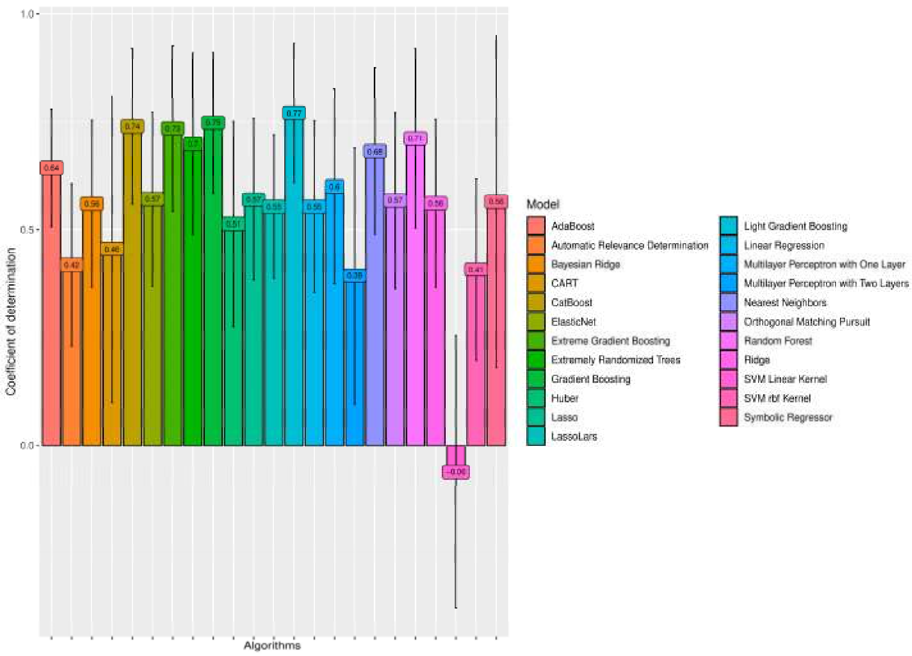

Figure 1. Coefficient of determination (R^2) of the 23 models fitted in default mode.

After examining the stacking mode of the Light Gradient Boosting, Nearest Neighbors, and Linear Regression algorithms, none of the nine arrangements showed improved predictive ability. The maximum R^2 achieved was 0.70 ± 0.04, while the minimum root mean squared error (RMSE) was 19.03 ± 2.23 and that of mean absolute error (MAE) was 3.18 ± 0.18. Similarly, after adjusting the stacking mode with four layers, which equates to 16 arrangements, the maximum R^2 obtained was 0.72 ± 0.03. The minimum RMSE achieved was 17.27 ± 2.14, while the minimum MAE was 2.99 ± 0.18. Although voting mode learning showed better performance than the three stacking variations (Table 4), it was unable to outperform Light Gradient Boosting in the default mode.

Table 4. Prediction errors in terms of the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) generated by the models in the voting mode.

Model	R^2 (Mean ± Std)	RMSE (Mean ± Std) [m^3 h^{-1}]	MAE (Mean ± Std)
Random Forest × Light Gradient Boosting	0.74 ± 0.03	16.20 ± 2.00	2.90 ± 0.16
Light Gradient Boosting × Nearest Neighbors	0.74 ± 0.03	16.20 ± 2.00	2.90 ± 0.16
Nearest Neighbors × Random Forest	0.74 ± 0.03	16.20 ± 2.00	2.90 ± 0.16
Nearest Neighbors × Light Gradient Boosting × Random Forest	0.75 ± 0.03	15.34 ± 2.02	2.81 ± 0.17

Even after utilizing different learning methods and adjusting hyperparameters, the Light Gradient Boosting algorithm in default mode provided the best yield prediction performance (Figure 2).

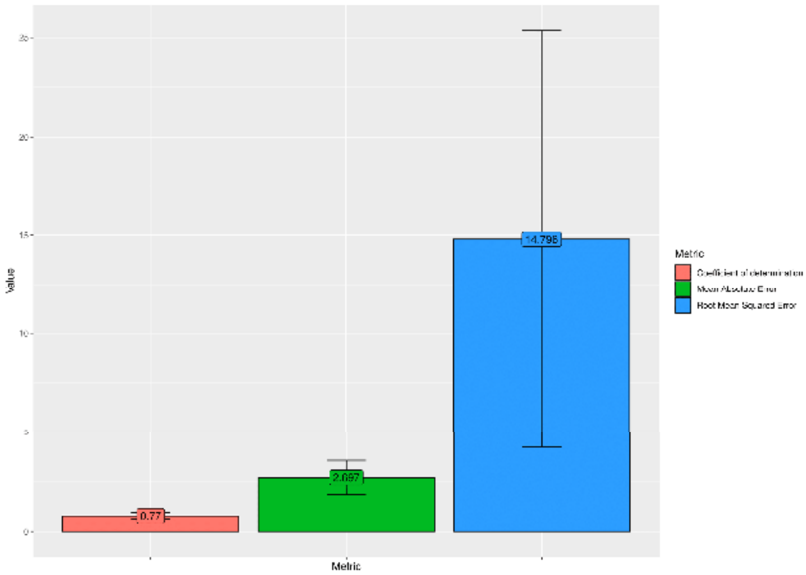

Figure 2. Quantitative metrics of the performance of the selected Light Gradient Boosting model.

4. Discussion

Each algorithm has a methodological basis that confers strengths and weaknesses that affect the performance of the model in response to solving the problem of predicting the productivity of forwarders. The default mode performance of each model, when comparing their results with the actual yield values, ranged from −0.06 for SVM Linear Kernel to 0.77 for Light Gradient Boosting. In addition to the size of the dataset, the results are also related to the structure of the dataset. Therefore, according to Goliatt et al. [51] and Altin et al. [52], it is fundamental to choose the correct algorithm for a specific problem.

Several factors influenced the differentiation pattern among the algorithms. As observed by Cipolla and Gondzio [53], the performance of SVM Linear Kernel is related to the high storage and manipulation requirements of dense and unstructured kernel matrices. The predictability performance has been improved by Light Gradient Boosting, which was developed to reduce memory consumption [54]. Sun et al.'s highly accurate prediction study [55] confirmed that the performance of Light Gradient Boosting is more robust compared to that of SVM. As a result, we found that boosting-based algorithms were more efficient in predicting forwarder productivity.

In addition to the performance of the algorithms, the results of the residual analysis, i.e., the RMSE and MAE, support our choice of the best fit model. In this analysis, even with a penalty, Light Gradient Boosting stood out among the other algorithms tested. According to Saaidi et al. [56], these fits are ensured by an iterative optimization process to minimize pseudoresidues. However, in addition to confirming the fit of the model by the R^2, we relate the performance of the model to the minimization of the error function between the actual productivity values and the predicted productivity values.

The integration of hyperparameter tuning with the learning strategy aims to enhance the predictive performance of the models generated. A comparison between the predicted yield values by Light Gradient Boosting and those obtained after the tuning process showed a reduction of 0.03 in the R^2 and increments of 1.83 and 0.17 in the RMSE and MAE, respectively. Likewise, the application of the combined voting and stacking learning method with three and four layers of algorithms resulted in a slight decline in the model performance. Nonetheless, under the default mode, Light Gradient Boosting persisted as the best performer.

Li et al. [57] suggested that the use of an integrated learning strategy should be limited to prediction problems with multiple stages, where different factors influence each stage. To predict the productivity of forwarders, the Light Gradient Boosting algorithm in default mode was sufficient, given that the influencing factors were generalized across all stages, and it ensured practicality and comprehensive predictive capability.

Tiernan et al. [58] observed that using influential factors to predict forwarder productivity in timber forwarding provides a convenient and transparent means for negotiating forest contracts and cost operations. Proposing expected forwarder productivity in tactical planning with an explanation of predictor variable variation has increased the autonomy of forest managers and, importantly, reduced unnecessary data waste from traditional approaches like time studies.

The developed machine learning model can assist forest managers in their activities by providing accurate estimates of expected timber volume, achieving an R^2 of 0.77. To acquire a more accurate estimate, forest managers can optimize harvesting logistics and tactically plan operations in accordance with production projections.

Malinovski et al. [59] found that the productivity of forwarders is influenced by transportation distance due to the need for possible machine trips. Cadei et al. [60] confirmed that the productivity of a forwarder is significantly affected by the distance traveled, load volume, and slope of the ground during movement. Taking these sources of uncertainty into account when modeling the productivity of a forwarder gives the forest manager confidence in scheduling operations, allocating teams and machinery, and defining routes in a timely manner based on informed decisions.

The utilization of data-driven resources has revolutionized forest management into intelligent forest management. The outcomes of the events and actions executed during log handling are indicators of logger productivity. Nevertheless, a data-driven process cannot exclusively depend on the creation of a machine learning model to maintain inference. Observation is a crucial subsequent step in incorporating and utilizing this resource in the tactical planning of logging operations.

Feng et al. [61] reported the development of several libraries in various programming languages that could serve this purpose. Klass et al. [62] reinforce that the reported probabilities from machine learning models can be misleading if the variables are not agnostically monitored.

Because of the constantly changing nature of the forest sector, the model was trained on a transient database. As a result, forest managers can evaluate the model by developing methods to demonstrate its ability to detect even minor deviations in monitoring, including technical and business dimensions, and comparing it to recent harvests.

5. Conclusions

The productivity of forwarders in timber forwarding in *Pinus* and *Eucalyptus* planted forests can be modeled using data-driven approaches. This modelling helps forest managers with tactical planning.

The default mode of the iterative optimization to minimize pseudoresidues using Light Gradient Boosting achieves outstanding performance in terms of productivity estimation, with a coefficient of determination of 0.77.

The factors affecting forwarder productivity, such as timber transport distance, terrain slope, plot area, species, and working shift, are generalized throughout the model construction process without employing built-in learning methods.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/f14091782/s1, Figure S1: Continuous variable data profile for the forwarder productivity predictive model.

Author Contributions: Conceptualization, R.A.M., D.S. and J.W.; methodology, R.A.M., D.S., R.O.A. and D.A.C.; software, R.A.M. and R.O.A.; validation, R.A.M., R.O.A., D.A.C., R.B.G.d.S., J.W. and D.S.; formal analysis, R.A.M., R.O.A. and D.A.C.; investigation, R.A.M., D.A.C. and R.B.G.d.S.; data curation, R.A.M. and R.O.A.; writing—original draft preparation, R.A.M., D.A.C. and R.B.G.d.S.; writing—review and editing, R.A.M., R.O.A., D.A.C., R.B.G.d.S., J.W. and D.S.; visualization; supervision, D.S. and J.W.; project administration, D.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: All data are provided in the main manuscript. Contact the corresponding author if further explanation is required.

Acknowledgments: We thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the financial support to the first author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rossi, V.; Claeys, F.; Bastin, D.; Gourlet-Fleury, S.; Guizol, P.; Eba'a-Atyi, R.; Sonwa, D.J.; Lescuyer, G.; Picard, N. Could REDD+ mechanisms induce logging companies to reduce forest degradation in Central Africa? *J. For. Econ.* **2017**, *29*, 107–117. [CrossRef]
2. Landhäusser, S.M. Impact of slash removal, drag scarification, and mounding on lodgepole pine cone distribution and seedling regeneration after cut-to-length harvesting on high elevation sites. *For. Ecol. Manag.* **2009**, *258*, 43–49. [CrossRef]
3. Prinz, R.; Spinelli, R.; Magagnotti, N.; Routa, J.; Asikainen, A. Modifying the settings of CTL timber harvesting machines to reduce fuel consumption and CO_2 emissions. *J. Clean. Prod.* **2018**, *197*, 208–217. [CrossRef]
4. 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]
5. Cudzik, A.; Brennensthul, M.; Białczyk, W.; Czarnecki, J. Tractive performance of tyres in forest conditions–Impact assessment of ground and tyres parameters. *Croat. J. For. Eng.* **2018**, *39*, 85–96.
6. Norihiro, J.; Ackerman, P.; Spong, B.D.; Längin, D. Productivity model for cut-to-length harvester operation in South African *eucalyptus* pulpwood plantations. *Croat. J. For. Eng.* **2018**, *39*, 1–13.
7. Obi, O.F.; Visser, R. Including exogenous factors in the evaluation of harvesting crew technical efficiency using a multi-step data envelopment analysis procedure. *Croat. J. For. Eng.* **2018**, *39*, 153–162.
8. Martell, D.L.; Gunn, E.A.; Weintraub, A. Forest management challenges for operational researchers. *Eur. J. Oper. Res.* **1998**, *104*, 1–17. [CrossRef]
9. Andersson, D.; Eriksson, L.O. Effects of temporal aggregation in integrated strategic/tactical and strategic forest planning. *For. Policy Econ.* **2007**, *9*, 965–981. [CrossRef]
10. Lin, T.; Rodríguez, L.F.; Shastri, Y.N.; Hansen, A.C.; Ting, K.C. Integrated strategic and tactical biomass–biofuel supply chain optimization. *Bioresour. Technol.* **2014**, *156*, 256–266. [CrossRef]
11. Fuentealba, S.; Pradenas, L.; Linfati, R.; Ferland, J.A. Forest harvest and sawmills: An integrated tactical planning model. *Comput. Electron. Agric.* **2019**, *156*, 275–281. [CrossRef]
12. Nasiri, S.; Khosravani, M.R. Applications of data-driven approaches in prediction of fatigue and fracture. *Mater. Today Commun.* **2022**, *33*, 104437. [CrossRef]
13. Wan, X.; Doherty, J. A data-driven approach for forecasting embankment settlement accounting for multi-stage construction. *Comput. Geotech.* **2022**, *152*, 105001. [CrossRef]

14. Marenić, E.; Seychal, G.; Passieux, J.-C. Data driven approach in multiphysics framework: Application to coupled electro-mechanical problems. *Comput. Methods Appl. Mech. Eng.* **2022**, *395*, 114959. [CrossRef]
15. Hasti, V.R.; Navarkar, A.; Gore, J.P. A data-driven approach using machine learning for early detection of the lean blowout. *Energy AI* **2021**, *5*, 100099. [CrossRef]
16. Tinmitondé, S.; He, X.; Yan, L.; Hounye, A.H. Data-driven prediction of critical flutter velocity of long-span suspension bridges using a probabilistic machine learning approach. *Comput. Struct.* **2023**, *280*, 107002. [CrossRef]
17. Meng, J.; Zhou, Y.-J.; Ye, T.-R.; Xiao, Y.-T.; Lu, Y.-Q.; Zheng, A.-W.; Liang, B. Hybrid data-driven framework for shale gas production performance analysis via game theory, machine learning, and optimization approaches. *Pet. Sci.* **2023**, *20*, 277–294. [CrossRef]
18. Sunhare, P.; Chowdhary, R.R.; Chattopadhyay, M.K. Internet of things and data mining: An application oriented survey. *J. King Saud Univ.-Comput. Inf. Sci.* **2020**, *34*, 3569–3590. [CrossRef]
19. Ma, S.; Fildes, R. Forecasting third-party mobile payments with implications for customer flow prediction. *Int. J. Forecast.* **2020**, *36*, 739–760. [CrossRef]
20. Sandoval Orozco, A.L.; Quinto Huamán, C.; Povedano Álvarez, D.; García Villalba, L.J. A machine learning forensics technique to detect post-processing in digital videos. *Futur. Gener. Comput. Syst.* **2020**, *111*, 199–212. [CrossRef]
21. Lim, P.K.; Julca, I.; Mutwil, M. Redesigning plant specialized metabolism with supervised machine learning using publicly available reactome data. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 1639–1650. [CrossRef]
22. Singh, P. Systematic review of data-centric approaches in artificial intelligence and machine learning. *Data Sci. Manag.* **2023**, *6*, 144–157. [CrossRef]
23. Eriksson, M.; Lindroos, O. Productivity of harvesters and forwarders in CTL operations in northern Sweden based on large follow-up datasets. *Int. J. For. Eng.* **2014**, *25*, 179–200. [CrossRef]
24. Guerra, S.P.S.; Oguri, G.; Spinelli, R. Harvesting *eucalyptus* energy plantations in Brazil with a modified New Holland forage harvester. *Biomass Bioenergy* **2016**, *86*, 21–27. [CrossRef]
25. Ghaffariyan, M.R.; Acuna, M.; Brown, M. Machine Productivity Evaluation for Harvesters and Forwarders in Thinning Operations in Australia. *Silva Balc.* **2019**, *20*, 13–25. [CrossRef]
26. Peng, W.; Karimi Sadaghiani, O. Enhancement of quality and quantity of woody biomass produced in forests using machine learning algorithms. *Biomass Bioenergy* **2023**, *175*, 106884. [CrossRef]
27. Speight, J.G.; Isbell, R.F. Soil Profiles. In *Australian Soil and Land Survey Field Handbook*; CSIRO: Canberra, Australia, 2009; pp. 132–133, ISBN 9780643093959.
28. Sotomayor, L.N.; Cracknell, M.J.; Musk, R. Supervised machine learning for predicting and interpreting dynamic drivers of plantation forest productivity in northern Tasmania, Australia. *Comput. Electron. Agric.* **2023**, *209*, 107804. [CrossRef]
29. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Computing Statistical: Vienna, Austria, 2022; ISBN 3-900051-07-0.
30. Naruei, I.; Keynia, F.; Sabbagh Molahosseini, A. Hunter–prey optimization: Algorithm and applications. *Soft Comput.* **2022**, *26*, 1279–1314. [CrossRef]
31. Obaidullah, A.J. Implementing and tuning machine learning-based models for description of solubility variations of nanomedicine in supercritical solvent for development of green processing. *Case Stud. Therm. Eng.* **2023**, *49*, 103200. [CrossRef]
32. 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. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830. [CrossRef]
33. Dong, M.; Sun, M.; Song, D.; Huang, L.; Yang, J.; Joo, Y.H. Real-time detection of wind power abnormal data based on semi-supervised learning Robust Random Cut Forest. *Energy* **2022**, *257*, 124761. [CrossRef]
34. Amati, M.; Tiede, J.; Sun, Q.; Deilami, K.; Hurley, J.; Fox, A.; Dickson, J. Using machine learning to identify urban forest crown bounding boxes (CBB): Exploring a new method to develop urban forest policy. *Urban For. Urban Green.* **2023**, *85*, 127943. [CrossRef]
35. Xie, Y.; Zhu, C.; Zhou, W.; Li, Z.; Liu, X.; Tu, M. Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances. *J. Pet. Sci. Eng.* **2018**, *160*, 182–193. [CrossRef]
36. Jamaleddyn, I.; El ayachi, R.; Biniz, M. An improved approach to Arabic news classification based on hyperparameter tuning of machine learning algorithms. *J. Eng. Res.* **2023**, *11*, 100061. [CrossRef]
37. Kim, S.; Yoon, H.C.; Lim, J.-T.; Jeong, D.; Kim, K.H. Productivity prediction in the Wolfcamp A and B using weighted voting ensemble machine learning method. *Gas Sci. Eng.* **2023**, *111*, 204916. [CrossRef]
38. Ghasemieh, A.; Lloyed, A.; Bahrami, P.; Vajar, P.; Kashef, R. A novel machine learning model with Stacking Ensemble Learner for predicting emergency readmission of heart-disease patients. *Decis. Anal. J.* **2023**, *7*, 100242. [CrossRef]
39. Nadirgil, O. Carbon price prediction using multiple hybrid machine learning models optimized by genetic algorithm. *J. Environ. Manag.* **2023**, *342*, 118061. [CrossRef]
40. Tavana, P.; Akraminia, M.; Koochari, A.; Bagherifard, A. An efficient ensemble method for detecting spinal curvature type using deep transfer learning and soft voting classifier. *Expert Syst. Appl.* **2023**, *213*, 119290. [CrossRef]
41. Shafighfard, T.; Bagherzadeh, F.; Rizi, R.A.; Yoo, D.-Y. Data-driven compressive strength prediction of steel fiber reinforced concrete (SFRC) subjected to elevated temperatures using stacked machine learning algorithms. *J. Mater. Res. Technol.* **2022**, *21*, 3777–3794. [CrossRef]

42. Arboretti, R.; Ceccato, R.; Pegoraro, L.; Salmaso, L. Design of Experiments and machine learning for product innovation: A systematic literature review. *Qual. Reliab. Eng. Int.* **2022**, *38*, 1131–1156. [CrossRef]
43. de Souza, L.P.B.; Ferreira, G.R.B.; Camerini, I.G.; de Correia, T.M.; de Ribeiro, M.G.C.; Hidalgo, J.A.S.; de São João, B.L.D.; Llerena, R.W.A.; Kubrusly, A.C.; Ayala, H.V.H.; et al. Machine learning-based cement integrity evaluation with a through-tubing logging experimental setup. *Geoenergy Sci. Eng.* **2023**, *227*, 211882. [CrossRef]
44. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. Catboost: Unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* **2018**, *2018*, 6638–6648.
45. Chen, T.; Guestrin, C. XGBoost. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; ACM: New York, NY, USA, 2016; pp. 785–794.
46. Begum, M.Y. Advanced modeling based on machine learning for evaluation of drug nanoparticle preparation via green technology: Theoretical assessment of solubility variations. *Case Stud. Therm. Eng.* **2023**, *45*, 103029. [CrossRef]
47. Brugere, L.; Kwon, Y.; Frazier, A.E.; Kedron, P. Improved prediction of tree species richness and interpretability of environmental drivers using a machine learning approach. *For. Ecol. Manag.* **2023**, *539*, 120972. [CrossRef]
48. Jayasinghe, T.; wei Chen, B.; Zhang, Z.; Meng, X.; Li, Y.; Gunawardena, T.; Mangalathu, S.; Mendis, P. Data-driven shear strength predictions of recycled aggregate concrete beams with/without shear reinforcement by applying machine learning approaches. *Constr. Build. Mater.* **2023**, *387*, 131604. [CrossRef]
49. Liemohn, M.W.; Shane, A.D.; Azari, A.R.; Petersen, A.K.; Swiger, B.M.; Mukhopadhyay, A. RMSE is not enough: Guidelines to robust data-model comparisons for magnetospheric physics. *J. Atmos. Solar-Terrestrial Phys.* **2021**, *218*, 105624. [CrossRef]
50. Sathishkumar, V.E.; Ramu, A.G.; Cho, J. Machine learning algorithms to predict the catalytic reduction performance of eco-toxic nitrophenols and azo dyes contaminants (Invited Article). *Alex. Eng. J.* **2023**, *72*, 673–693. [CrossRef]
51. Goliatt, L.; Saporetti, C.M.; Oliveira, L.C.; Pereira, E. Performance of evolutionary optimized machine learning for modeling total organic carbon in core samples of shale gas fields. *Petroleum*, 2023; *in press*. [CrossRef]
52. Altin, F.G.; Budak, İ.; Özcan, F. Predicting the amount of medical waste using kernel-based SVM and deep learning methods for a private hospital in Turkey. *Sustain. Chem. Pharm.* **2023**, *33*, 101060. [CrossRef]
53. Cipolla, S.; Gondzio, J. Training very large scale nonlinear SVMs using Alternating Direction Method of Multipliers coupled with the Hierarchically Semi-Separable kernel approximations. *EURO J. Comput. Optim.* **2022**, *10*, 100046. [CrossRef]
54. Thongthammachart, T.; Araki, S.; Shimadera, H.; Matsuo, T.; Kondo, A. Incorporating Light Gradient Boosting Machine to land use regression model for estimating NO2 and PM2.5 levels in Kansai region, Japan. *Environ. Model. Softw.* **2022**, *155*, 105447. [CrossRef]
55. Sun, X.; Liu, M.; Sima, Z. A novel cryptocurrency price trend forecasting model based on LightGBM. *Financ. Res. Lett.* **2020**, *32*, 101084. [CrossRef]
56. Saaidi, A.; Bichri, A.; Abderafi, S. Efficient machine learning model to predict dynamic viscosity in phosphoric acid production. *Results Eng.* **2023**, *18*, 101024. [CrossRef]
57. Li, N.; Wu, Y.; Wang, Q.; Ye, H.; Wang, L.; Jia, M.; Zhao, S. Underground mine truck travel time prediction based on stacking integrated learning. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105873. [CrossRef]
58. Tiernan, D.; Zeleke, G.; Owende, P.M.; Kanali, C.; Lyons, J.; Ward, S. Effect of Working Conditions on Forwarder Productivity in Cut-to-length Timber Harvesting on Sensitive Forest Sites in Ireland. *Biosyst. Eng.* **2004**, *87*, 167–177. [CrossRef]
59. Malinovski, R.A.; Malinovski, R.A.; Malinovski, J.R. Análise das variáveis de influência na produtividade das máquinas de colheita de madeira em função das características físicas do terreno, do povoamento e do planejamento operacional florestal. *Floresta* **2006**, *36*, 169–182. [CrossRef]
60. Cadei, A.; Mologni, O.; Röser, D.; Cavalli, R.; Grigolato, S. Forwarder Productivity in Salvage Logging Operations in Difficult Terrain. *Forests* **2020**, *11*, 341. [CrossRef]
61. Feng, X.; Wang, D.; Hou, B.; Yan, T. Interpretable federated learning for machine condition monitoring: Interpretable average global model as a fault feature library. *Eng. Appl. Artif. Intell.* **2023**, *124*, 106632. [CrossRef]
62. Klass, L.; Kabza, A.; Sehnke, F.; Strecker, K.; Hölzle, M. Lifelong performance monitoring of PEM fuel cells using machine learning models. *J. Power Sources* **2023**, *580*, 233308. [CrossRef]