Innovaciencia ISSN: 2346-075X Bucaramanga - Colombia

Short communications

Innovaciencia 2026; 14(1): e5883

Data mining models to predict timber production across Colombian departments

Modelos de minería de datos para predecir la producción de madera en los departamentos de Colombia

Juan Jacobo Galindo Suarez[1,2*], Víctor Hugo Aristizábal[3], Fredy Angarita Reina[4], Francisco Javier Vélez Hoyos[3]

How to cite this article: Galindo Suarez JJ, Aristizabal VH, Angarita Reina F, Vélez Hoyos FJ. Data mining models to predict timber production across Colombian departments. Innovaciencia. 2026;14(1):e5883. https://doi.org/10.15649/2346075X.5883

Received: 12 November 2025
Accepted: 03 June 2026
Published: 23 June 2026

Highlights

- Open ICA data and data mining models enabled four-quarter forecasts of departmental timber mobilization in Colombia.
- ARIMA and Random Forest showed the best predictive performance depending on the stability or variability of each departmental time series.
- The CRISP-DM workflow with KSSA imputation provided a reproducible approach for analyzing timber mobilization trends and supporting forest-sector planning.

[1] Universidad Cooperativa de Colombia, Medellín, Colombia.
[2] Universidad Icesi, Cali, Colombia.
[3] Facultad de Ingeniería, Grupo de Investigación Termomec, Universidad Cooperativa de Colombia, Medellín, Colombia.
[4] Facultad de Ingeniería, Universidad Cooperativa de Colombia, Bucaramanga, Colombia.
*Corresponding Author: juanjacobogalindosuarez@gmail.com

Innovaciencia ISSN: 2346-075X Bucaramanga - Colombia

Short communications

ABSTRACT

Introduction. Timber production in Colombia is strategic for economic development and environmental conservation, yet reliable predictive tools remain scarce. **Objective.** To evaluate the performance of statistical and machine-learning models for forecasting department-level timber mobilization volumes in Colombia using open data from the Colombian Agricultural Institute (2012–2022). **Materials and Methods.** Following the CRISP-DM framework, we performed data cleaning and preprocessing, imputed missing values via KSSA, and implemented five model families (ARIMA, Prophet, GLMNET, Random Forest, and Prophet Boost). Models were trained on 90% of the historical series and evaluated with RMSE, MAE, and MAPE. **Results.** ARIMA and Random Forest achieved the best performance depending on the stability or variability of each series, enabling reliable four-quarter-ahead forecasts. Departments such as Antioquia, Valle del Cauca, and Cauca are projected to maintain high production levels, whereas Meta and Casanare exhibit greater instability. **Conclusions.** These findings underscore the value of integrating open data and machine-learning techniques to support the sustainable management of Colombia's forest resources.

Keywords. Timber production; Data mining; Machine learning; Colombia; Forecasting; Time series.

RESUMEN

Introducción. La producción de madera en Colombia es estratégica para el desarrollo económico y la conservación ambiental, pero aún carece de herramientas predictivas confiables. **Objetivo.** Evaluar el desempeño de modelos estadísticos y de aprendizaje automático para el pronóstico del volumen de madera movilizada a nivel departamental en Colombia utilizando datos abiertos del Instituto Colombiano Agropecuario (ICA) correspondientes al período 2012–2022. **Materiales y Métodos.** El análisis siguió la metodología CRISP-DM, con limpieza y tratamiento de datos, imputación mediante KSSA y la implementación de cinco familias de modelos (ARIMA, Prophet, GLMNET, Random Forest y Prophet Boost). Los modelos se entrenaron con el 90 % de la serie histórica y se evaluaron con métricas como RMSE, MAE y MAPE. **Resultados.** ARIMA y Random Forest alcanzaron el mejor desempeño según la estabilidad o variabilidad de las series, generando pronósticos confiables a cuatro trimestres. Departamentos como Antioquia, Valle del Cauca y Cauca mantendrían altos niveles de producción, mientras que Meta y Casanare presentarían mayor inestabilidad. **Conclusiones.** Los hallazgos confirman el valor de integrar datos abiertos y técnicas de aprendizaje automático para la gestión sostenible del recurso forestal en Colombia.

Palabras Clave. Producción de madera; Data mining; Machine learning; Colombia; Pronóstico; Series de tiempo.

Innovaciencia ISSN: 2346-075X Bucaramanga - Colombia
Short communications

INTRODUCTION

Timber production in Colombia represents a strategic activity from both an economic and environmental perspective. In recent years, approximately 76% of the mobilized volume has come from commercial forest plantations[1], located primarily in departments such as Antioquia, Nariño, Cauca, Valle del Cauca, and Chocó [2]. However, a considerable portion of the national supply still depends on natural ecosystems, and approximately 10% of the country's deforestation is associated with the illegal timber market [3,4]. This situation places significant pressure on the country's forests and biodiversity, affecting key ecosystem services [5]. In this regard, having reliable information on current and future timber production is essential for designing policies aimed at the sustainable use of forests and the conservation of biodiversity.

Globally, approximately 3.4 billion cubic meters of timber are produced annually since 2005 [5,6]. The magnitude of this activity highlights its economic importance, but also the need for adequate management tools that allow balancing production with ecological sustainability. Countries such as Ecuador have applied data mining methodologies to agricultural crops such as cacao, obtaining predictive models that estimate future production and guide planning [7]. Similarly, in China a model based on PSO-SVM was developed that achieved an average forecast error of only 4.017%, validating the usefulness of these techniques for anticipating forest production dynamics [8]. More recently, in Brazil, approaches such as Random Forest and ANFIS were evaluated for predicting forest production, with highly accurate results [9]. These precedents demonstrate that the use of data analytics models is a robust and replicable approach for anticipating scenarios in the forestry sector.

In the Colombian context, although recent figures on production and areas of forest plantations exist, official predictive models that allow estimating the volume of timber mobilized in the coming years are still unavailable [2]. This absence of tools limits the capacity of institutions to plan conservation policies, reduce pressure on natural forests, and combat the illegal timber trade [3]. Consequently, it is necessary to implement analytical models that integrate open data and data mining techniques, capable of generating reliable forecasts and supporting the sustainable management of timber resources.

Within this framework, concepts such as commercial forest plantations, understood as crops of fast-growing or high-commercial-value forest species for productive purposes, play a central role, as they represent the core of the country's timber production [2]. To anticipate the dynamics of these plantations, it is necessary to apply data analytics,

understood as the science that examines raw information with the purpose of obtaining useful conclusions for decision-making [10]. Likewise, the use of predictive models based on time series and machine learning allows leveraging historical information to anticipate production scenarios and support planning decisions. In particular, the comparison of classical statistical approaches with non-linear algorithms allows evaluating which model families best adapt to series with stability or high territorial-level variability, contributing to the prediction of complex environmental and productive processes.

To systematically structure this process, the research is framed within the CRISP-DM (Cross-Industry Standard Process for Data Mining), which organizes the phases of a data mining project from problem understanding to the deployment of predictive models [11]. This standard, widely used in different fields, provides a systematic framework for data preparation, algorithm training, and model evaluation, ensuring the replicability and quality of the results obtained.

Accordingly, this work proposes the design and implementation of a data analytics model for predicting the volume of timber mobilized in the departments of Colombia, using open records from the Colombian Agricultural Institute (ICA) between 2012 and 2022. Classical and modern time series prediction algorithms are employed, including ARIMA, Prophet, GLMNET, Random Forest, and Prophet Boost. ARIMA and Prophet are based on statistical assumptions and additive structures; GLMNET is a regularized regression model over temporal features; Random Forest is a non-linear ensemble-based method; and Prophet Boost integrates additive components with boosting techniques. These models differ in their stationarity assumptions, ability to capture non-linearities, and interpretability, which allows evaluating their suitability under different departmental series behaviors, complemented by robust imputation techniques such as KSSA. The contribution of this study lies not only in building a functional and reproducible model, but also in generating scientific evidence to support strategic decision-making in the Colombian forestry sector, with direct implications for productive planning, biodiversity conservation, and the fulfillment of the Sustainable Development Goals.

MATERIALS AND METHODS

The work was structured under the CRISP-DM methodology, which organized the analytical cycle into six phases (business understanding, data understanding, preparation, modeling, evaluation and deployment), serving as an

operational guide for technical and documentation decisions. The main source was the open database of the Colombian Government titled "Database related to timber mobilized from Commercial Forest Plantations" [12], with metadata created on 24-Nov-2022 and updated on 26-Mar-2024, containing quarterly records by department for 2012–2022 and key variables: Year, TRIMESTRE, DPTO, and VOLUMEN (m^3). Processing was performed in R/RStudio 4.4.1 and automated via scripts to ensure reproducibility with free software. The project organization included dedicated directories for images (maps and figures), tables (intermediate and final outputs), series (imputation traces), maps (annual cartography), and select (metric diagnostics), consistent with the CRISP-DM logic.

The analysis was performed using the variables Year, DPTO, TRIMESTRE, and VOLUMEN (m^3). To build the temporal dimension, a Date variable was generated by combining year-quarter and assigning the first day of each quarter (Q1→YYYY-01-01; Q2→YYYY-04-01; Q3→YYYY-07-01; Q4→YYYY-10-01), ensuring homogeneous quarterly periodicity by department. The volume average per DPTO and Date was then calculated, and a complete grid of all Date (2012–2022, quarterly) × DPTO combinations was created, which allowed aligning missing values and standardizing imputation and modeling. The amount of data per department was evaluated and those with more than 30% missing values were excluded. For cases with incomplete but acceptable data (≤30% NAs), KSSA (k-nearest Singular Spectrum Analysis) was applied with a multi-method search (including auto.arima, StructTS, linear smoothing, and simple and linear moving averages)[13]. With three segments and automatic selection of the best reconstruction, missing values were replaced and the resulting series were stored for traceability and quality control.

For the development of the predictive model, a department-level workflow was implemented for the fitting, calibration, and comparison of five predictive model families: ARIMA (auto_arima), Prophet, GLMNET (L1/L2 penalized regression), Random Forest (ranger), and Prophet Boost (Prophet + XGBoost). The process was carried out using the tidymodels ecosystem in R, utilizing modeltime (temporal model management), parsnip (model declaration), recipes (preprocessing) and workflows (integration of recipes and estimators). For each department, data partitioning was performed using a holdout temporal validation scheme (90% training, 10% testing), reserving the most recent segment of the series to evaluate predictive performance. This decision was based on the limited length of the quarterly series (approximately 40 observations), prioritizing a sufficiently large training set to capture the temporal structure while avoiding look-ahead bias.

Once the temporal partition was established, the models were calibrated on the test set using modeltime_calibrate, recording the metrics RMSE, MAE, MAPE, SMAPE, and MASE. The best model at the departmental level was identified and then refitted using 100% of the historical series to generate four-quarter-ahead forecasts. Results were consolidated in spreadsheets and comparative charts (observed vs. predicted) by department.

For the territorial analysis, administrative cartography (GADM, departmental level) and the annual average table by DPTO were integrated; after harmonizing place names, global deciles of "promedio_anual_volumen" were calculated and each department was categorized by year using these thresholds. Using ggplot2 and sf, annual maps (2012–2022) were produced and exported to PNG, complementing the performance report and communication of results. Script automation and standardized output naming facilitate workflow repeatability upon new data updates.

Data Availability. The dataset supporting the findings of this study has been deposited in Mendeley Data and is publicly accessible under the title: Galindo, Jacobo; Aristizábal, Victor Hugo; Angarita, Fredy; Vélez, Francisco (2026), "Data mining models to predict timber production across Colombian departments. - Dataset", Mendeley Data, V1,.1 http://doi.org/10.17632/7gjg9s77yp

RESULTS

Initial data description

The original dataset comprised records of the volume of timber mobilized by commercial forest plantations in the departments of Colombia between 2012 and 2022. The average quarterly volume ranged from below 1,000 m^3 to above 50,000 m^3, with departments such as Antioquia, Valle del Cauca, and Cauca standing out for their high timber mobilization levels. The exploratory analysis revealed missing values in several departmental series, particularly in departments with low forestry activity or irregular reporting. Departments with an absence greater than 30% of records were excluded from subsequent analysis. For departments with a sufficient number of observations (≤30% missing values), data imputation was applied using the KSSA method, selecting for each department the best-fitting model for its time series.

Spatio-temporal analysis results: Both consistent and fluctuating patterns were observed throughout the evaluation period, spanning from 2012 to 2022. In particular, departments such as Antioquia, Valle del Cauca, Cauca, Nariño, and Chocó consistently showed high average volumes in most years, positioning them as leaders in commercial forest

production. On the other hand, departments such as Huila, Tolima, Putumayo, Cundinamarca, Santander, Norte de Santander, and La Guajira maintained low mobilization levels, likely reflecting a combination of logistical constraints, low levels of commercial plantation establishment, or environmental conservation priorities. Finally, it is important to note that the departments of Bolívar, Córdoba, Boyacá, Chocó, Nariño, Caquetá, Amazonas, Vaupés, Guainía, and San Andrés y Providencia do not appear in the spatio-temporal results maps because they either have no data in the original database or are missing more than 30% of the data required to run predictions correctly.

In 2012, the first year recorded in the ICA database, Meta and Cesar were the departments with the highest timber production levels (Figure 1a). In 2013, the average timber production volume decreased compared with the previous year, whereas in 2014 it increased, with Antioquia, Valle del Cauca, and Vichada emerging as the leading production departments. In 2015, Cesar, Risaralda, and Chocó also ranked among the main producers. By 2016, Caldas and Casanare were included among the top-producing departments, and this year recorded the highest average timber production in the ICA database. In 2017, timber production declined, with Cauca and Valle del Cauca reporting the highest production levels. Antioquia again appeared among the leading producers in 2018, while Sucre, Caldas, and Risaralda were among the departments with the highest production in 2019. In 2020, production levels remained similar to those observed in the previous year. In 2021, Vichada was again identified among the top-producing departments, together with several of the departments previously mentioned. Finally, in 2022, Sucre and Casanare were also among the departments with the highest timber production. Representative timber production time series for the departments of Meta and Valle del Cauca are presented, while the complete set of departmental series is provided in Data Availability.

Predictive model results

Five models were fitted for each department (ARIMA, Prophet, GLMNET, Random Forest, and Prophet Boost). Model performance was assessed using a test set comprising the last 10% of observations and five evaluation metrics (RMSE, MAE, MAPE, SMAPE, and MASE).

The results were consolidated to identify the best-performing model for each metric and department. Prophet showed the best performance for Antioquia, whereas ARIMA performed best for Atlántico, Caldas, Cesar, Cundinamarca, Magdalena, Meta, Nariño, Putumayo, and Sucre. Ranger (Random Forest) was the best-performing model for

Casanare, Cauca, Huila, Norte de Santander, Quindío, Risaralda, Santander, Tolima, Valle del Cauca, and Vichada. GLMNET and Prophet Boost are not reported because of their poor performance. In addition, department-level results were graphically compiled, with a representative example shown in (**Figure 3**).

Forecasting results. Once the best-performing model for each department had been identified according to the evaluation metrics, four-quarter forecasts corresponding to a one-year horizon were generated (**Table 1**).

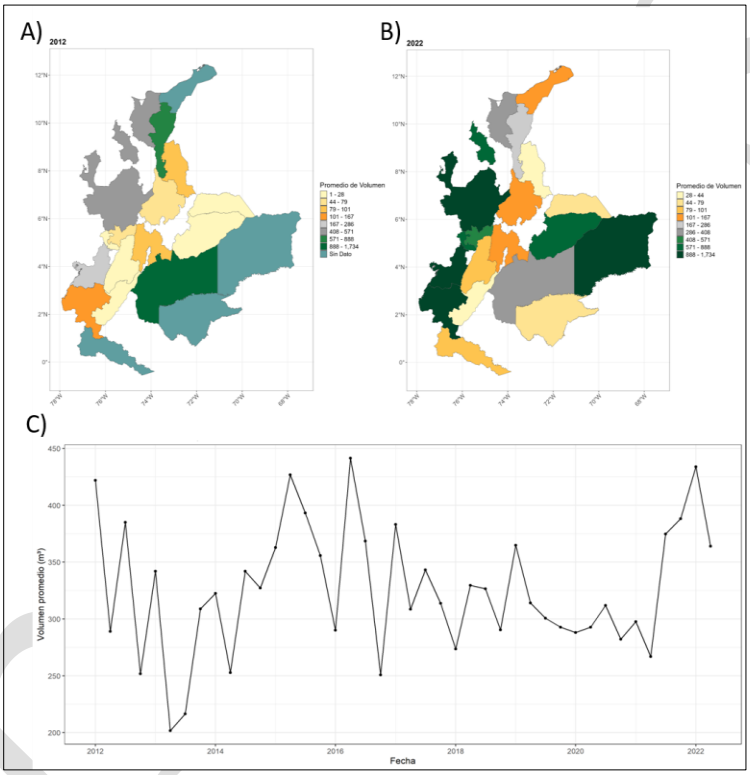

Figure 1. A) Decile map of the average timber volume produced per quarter across the different departments of Colombia from the initial year (2012) of the ICA database record**. B)** Decile map of the average timber volume produced per quarter across the different departments of Colombia through the final year (2022) of the ICA database record**. C)** Time series of the average volume produced in Colombia per quarter.

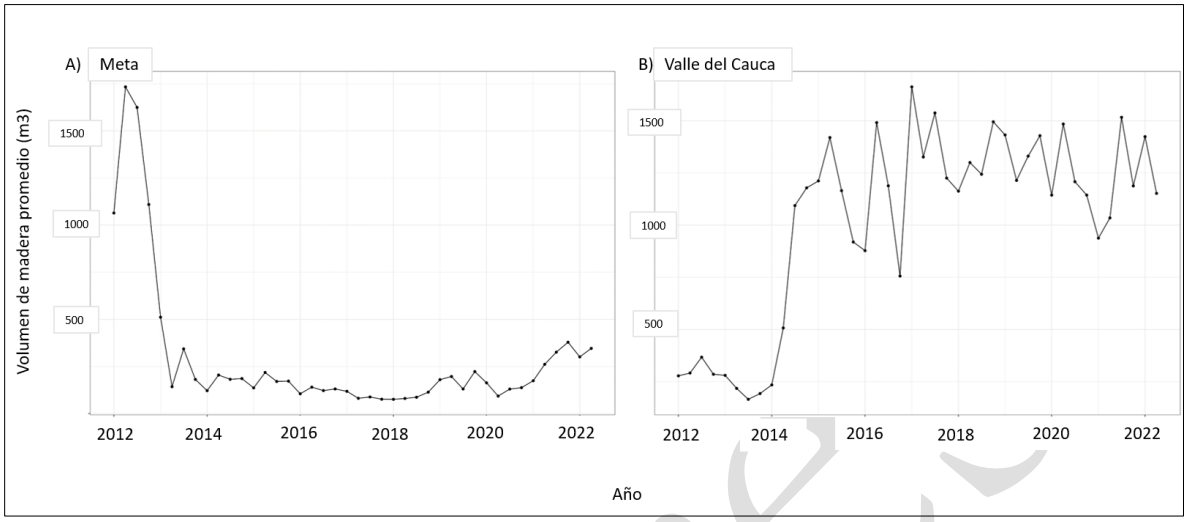

Figure 2. Representative examples of the timber production time series for the departments of **A)** Meta and **B)** Valle del Cauca

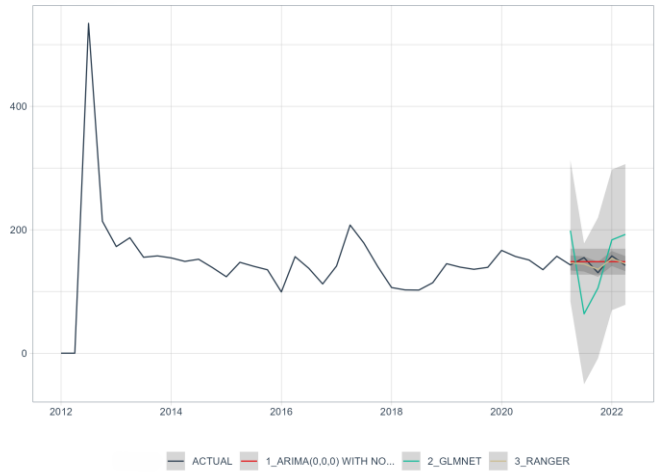

Figure 3. Evaluation of the prediction models according to the metrics (RMSE, MAE, MAPE, SMAPE, and MASE). This example shows one of the charts for the department of Boyacá, in which the prediction model best fitted to the predicted 10% was the Ranger (Random Forest) model.

Table 1. Final timber volume (m3) recorded in the ICA data table and predicted averages for each department over the next 4 quarters (one year).

Deparment	Latest Value	2022-04-01	2022-07-01	2022-10-01	2023-01-01
Antioquia	949.9227	737.1146	764.2778	721.5359	775.1024
Atlántico	125.8182	227.0496	217.5752	210.3996	228.0402
Caldas	498.4706	421.7245	448.0353	426.3642	420.3207
Casanare	520.931	274.1216	281.7329	344.4266	273.3596
Cauca	1407.976	1343.086	1396.452	1202.351	1174.811
Cesar	238.05	245.6914	244.5452	287.7523	256.6256
Cundinamarca	119.5	90.38343	89.44499	86.51218	94.80245
Huila	47.125	36.36366	42.02396	42.37249	37.10909
Magdalena	384.7059	250.0387	251.0172	240.6762	256.9698
Meta	345.5714	145.2857	142.1299	158.1792	164.2626
Nariño	61.94118	69.00144	73.78038	72.69873	67.63853
Norte De Santander	46.55556	13.29899	13.48179	14.36722	12.71836
Putumayo	28.33333	24.91079	37.33342	48.11121	29.10167
Quindío	230.1333	441.2671	443.8246	431.475	446.6453
Risaralda	693.4286	709.6308	709.0226	673.8424	634.6577
Santander	158.2245	99.20466	111.2449	111.322	100.5902
Sucre	657.6	388.3318	310.8443	290.1842	560.2706
Tolima	122.1481	170.0813	157.9781	155.9474	166.0716
Valle Del Cauca	1152.506	1196.517	1186.165	1170.222	1104.561
Vichada	948.4444	212.3194	160.8758	171.3598	264.9868

DISCUSSION

Previous research has demonstrated the applicability of data mining techniques for forecasting forest production dynamics, as evidenced by studies conducted in China [8], Ecuador [7], and Brazil [9]. This work developed a reproducible and reliable modeling framework capable of generating department-level forecasts over a four-quarter horizon and automatically selecting the best-performing algorithm for each territory according to error metrics such as RMSE, MAE, and MAPE. This approach validated that, as in the international studies consulted, the combination of classical techniques with advanced algorithms improves predictive power when adapted to specific contexts.

The results obtained showed that the best-performing models in most departments were ARIMA and Random Forest (Ranger), which aligns with their characteristics and the behavior of the evaluated time series. ARIMA stood out particularly in departments with stable temporal patterns and low variability. As a classical time series model, ARIMA makes very good use of linear autocorrelation information within historical series [14,15]. This behavior was observed in departments such as Atlántico, Caldas, Cesar, Meta, and Nariño, whose series exhibited continuity, smooth trends, or clear cycles. Residual diagnostics showed no significant autocorrelation, suggesting an adequate model fit. Likewise, stationarity tests supported the suitability of this approach for the analyzed series (see Data Availability). On the other hand, Random Forest, as a tree-based machine learning model, performed well in departments with greater variability, outliers, or abrupt changes. Random Forest does not require the series to be stationary and can better adapt to non-linear structures. This was evidenced in departments such as Cauca, Valle del Cauca, Córdoba, Quindío, Tolima, and Vichada, where the dispersion or irregularity in the data favored the use of a more flexible model (see Data Availability).

The GLMNET and Prophet Boost models showed inferior performance compared to ARIMA and Random Forest based on the evaluated metrics (see Data Availability). The lower performance of these models on the volume series can be attributed to the fact that GLMNET imposes penalties that are useful for regularization and variable selection, but is not especially effective in time series where the variable depends mainly on its past behavior rather than on multiple independent predictors [16]. Since the series were univariate, the model did not have enough "explanatory" features or predictor variables for its penalization structure to be meaningful. Likewise, Prophet Boost, although more powerful than classical Prophet, requires good parameterization to correctly detect seasonal changes and trends. In this case, many series were short and highly irregular, which limits the model's ability to identify robust seasonalities.

Furthermore, Prophet Boost is sensitive to overfitting when data are scarce, which may have affected its performance in departments with inconsistent records.

Several important limitations were identified during the study. The main limitation was data quality and continuity, as some departments had a high proportion of missing data (>30%), which required their exclusion from the predictive analysis. Another data-related limitation is that the records only extend through 2022 and the database has not been updated further. There are also difficulties in modeling extreme values; although KSSA imputation was effective, in some cases with abrupt production changes (e.g., economic or logistical shocks), the models tended to smooth the curves, reducing the ability to anticipate extreme events.

For future research, it is suggested to complement the ICA database with more recent data, as the available records currently extend only through 2022. In addition, data from other institutions should be incorporated, given the existence of informal markets not captured in the ICA database; this could increase data availability and support more accurate analyses. The inclusion of exogenous explanatory variables beyond time, such as temperature, precipitation, and land use, should also be considered. Finally, future studies could implement deep learning models, spatial prediction approaches, or vegetation indices, such as NDVI or LAI, to anticipate forest productivity.

These findings highlight the usefulness of combining open data, accessible algorithms, and free software to generate reproducible forecasting tools for departmental-level commercial forest production in Colombia. Rather than replacing existing environmental information systems, this approach may complement them by providing timely evidence on production trends and territorial variability. In this sense, the forecasts generated in this study may support decision-making in the productive and environmental sectors, particularly for anticipating supply scenarios, informing sectoral planning, and guiding forest resource management.

CONCLUSIONS

This study provides a structured application of data mining techniques for forecasting commercial forest production at the departmental level in Colombia. The findings suggest that the integration of open data, accessible algorithms, and free software can generate functional and reproducible predictive models with potential applications in environmental information systems. This approach supports the research hypothesis and provides a useful tool for anticipating production scenarios and informing timber resource management.

From a technical perspective, an analytical workflow based on the CRISP-DM methodology was consolidated. This included data cleaning, structuring, and missing-value imputation using KSSA, which enabled the construction of consistent time series for each department. Five forecasting algorithms were evaluated: ARIMA, Prophet, GLMNET, Random Forest, and Prophet Boost. Overall, ARIMA and Random Forest showed the best performance in most cases, depending on the stability or irregularity of the analyzed series. The four-quarter forecasts indicated that Antioquia, Valle del Cauca, and Cauca are expected to remain among the leading producers, whereas Meta, Casanare, and Chocó exhibited greater variability in their production trends.

From a strategic and environmental standpoint, the results provide relevant input for sectoral planning and evidence-based decision-making in forest resource management. Furthermore, the methodological framework developed in this study has the potential to be replicated for other natural resources, productive sectors, or regions.

ETHICAL CONSIDERATIONS

This study used the open database of the Colombian Government titled "Database related to timber mobilized from Commercial Forest Plantations", with metadata created on 24-Nov-2022 and updated on 26-Mar-2024. No human participants, animals, or personal data were involved; therefore, ethical approval and informed consent were not required

FUNDING

Funding was provided by the project "Implementation of actions for the protection of water basins and soils through reforestation with emerging technologies and biotechnology in the Eastern Plains region in the departments of Meta and Arauca", with code BPIN2022000100005, executed by the Universidad Cooperativa de Colombia, campuses Medellín, Villavicencio, and Arauca, and funded by the Ministry of Science and Technology (Minciencias) of Colombia through the Science, Technology, and Innovation Fund of the General Royalties System (SGR)

CONFLICT OF INTEREST

The authors declare no conflict of interest

REFERENCES

1. **Martínez-Cortés ÓG, Kant S, Isufllari H.** An analysis of wood availability under six policy scenarios of commercial forest plantations in Colombia. Forest Policy and Economics. 2022;138:102722. https://doi.org/10.1016/j.forpol.2022.102722
2. **Fedemaderas.** Boletín Estadístico Forestal 2022 – Federación Nacional de Industriales de la Madera [Internet]. 2022 [cited 2023 May 13]. Available from: https://fedemaderas.org.co/boletin-estadistico-forestal-2022/
3. **Fondo Mundial para la Naturaleza - WWF.** Madera legal: un mercado lleno de oportunidades para Colombia [Internet]. 2023 [cited 2023 Sep 22]. Available from: https://www.wwf.org.co/?380872/Madera-legal-un-mercado-lleno-de-oportunidades-para-Colombia
4. **Scientific Reports.** Deforestation in Colombian protected areas increased during post-conflict periods [Internet]. [cited 2026 Apr 20]. Available from: https://www.nature.com/articles/s41598-020-61861-y
5. **Verkerk PJ, Levers C, Kuemmerle T, Lindner M, Valbuena R, Verburg PH, et al.** Mapping wood production in European forests. Forest Ecology and Management. 2015;357:228–38. https://doi.org/10.1016/j.foreco.2015.08.007
6. **FAO.** Global Forest Resources Assessment 2020. Key findings [Internet]. 1st ed. Rome: FAO; 2020 [cited 2026 Apr 20]. Available from: https://openknowledge.fao.org/items/ac91b7b4-87eb-41eb-bdb1-d1c31fe249a8
7. **Mazon B, Jaramillo M, Romero O, Borja A, Aguirre M, Contento M.** Tecnologías de Inteligencia de Negocios y Minería de datos para el análisis de la producción y comercialización de cacao. Revista ESPACIOS [Internet]. 2018 [cited 2023 Jun 13];39(32). Available from: https://www.revistaespacios.com/a18v39n32/18393206.html
8. **Wu P, Yi X, Jin K.** A study on Chinese output of timber prediction model based on PSO-SVM. Advances in Information Sciences and Service Sciences. 2012;4(2):227–33. https://doi.org/10.4156/aiss.vol4.issue2.28
9. **Pereira Martins Silva J, Luiza Marques da Silva M, Ribeiro de Mendonça A, Fernandes da Silva G, Almeida de Barros Junior A, Ferreira da Silva E, et al.** Prognosis of forest production using machine learning techniques. Information Processing in Agriculture. 2023;10(1):71–84. https://doi.org/10.1016/j.inpa.2021.09.004
10. **Yasar K.** What is data analytics? | Definition from TechTarget [Internet]. Search Data Management. 2024 [cited 2025 Sep 16]. Available from: https://www.techtarget.com/searchdatamanagement/definition/data-analytics
11. **Espinosa-Zúñiga JJ.** Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública. Ingeniería, investigación y tecnología. 2020;21(1). https://doi.org/10.22201/fi.25940732e.2020.21n1.008
12. **Ministerio de Agricultura y Desarrollo Rural.** Base de datos relacionada con madera movilizada proveniente de Plantaciones Forestales Comerciales | Datos Abiertos Colombia [Internet]. 2024 [cited 2025 Sep 3]. Available from: https://www.datos.gov.co/Agricultura-y-Desarrollo-Rural/Base-de-datos-relacionada-con-madera-movilizada-pr/9aan-wm8m/about_data
13. **Mahmoudvand R, Rodrigues PC.** Missing value imputation in time series using Singular Spectrum Analysis. Int J Energy Stat. 2016;4(1):1650005. https://doi.org/10.1142/S2335680416500058
14. **Hyndman RJ, Athanasopoulos G.** Forecasting: Principles and Practice. 2nd ed [Internet]. Melbourne: OTexts; 2018 [cited 2025 Sep 3]. Available from: https://otexts.com/fpp2/
15. **Yadav S, Shukla S.** A comparative study of ARIMA, Prophet and LSTM for time series prediction. J Artif Intell Mach Learn Data Sci. 2022;1(1):1813–6. https://doi.org/10.51219/JAIMLD/sandeep-yadav/402

16. **Engebretsen S, Bohlin J.** Statistical predictions with glmnet. Clin Epigenetics. 2019;11:123. https://doi.org/10.1186/s13148-019-0730-1