

# Estrategias de gestión de la cadena productiva del café empleando analítica de datos

**Karen Julieth Castellanos Rodríguez<sup>1</sup>, Óscar Mayorga Torres<sup>2</sup>**

**<sup>1</sup>Universidad de Pamplona. Pamplona, Colombia**

**<sup>2</sup>Universidad Francisco de Paula Santander. Cúcuta, Colombia**

## Resumen

La presente investigación propone estrategias de gestión de la cadena productiva del café empleando analítica de datos, para la mejora en la toma de decisiones en la gestión del cultivo y como actor de análisis el productor (caficultor); la base estadística se conformó con datos de la producción, calidad, áreas cultivadas y cantidad de productores del grano en el departamento de Norte de Santander; la ventana de tiempo de los datos se estableció del 2013-2024, con la cual se construyó un análisis exploratorio permitiendo entender el comportamiento del sector caficultor en el departamento y los impactos en la productividad y competitividad en el ámbito nacional.

Con lo anterior, se establecieron los impactos que afectan la productividad como dispersión geográfica, condiciones de siembra, manipulación en la cosecha y postcosecha; con una pérdida acumulada al 2024 del 12%; lo que permitió hacer cruces relacionales entre diferentes denominaciones asociativas de los productores, del área cultivada, la disponible y la renovación o transferencia de cultivos; de lo anterior, y para dar sustento a las estrategias se emplearon técnicas de clustering y modelos de predicción avanzada, segmentando los cultivos por condiciones productivas, área cultivable y operación del cultivo.

Finalmente, el estudio muestra la importancia del sector caficultor para la economía de la región proponiendo estrategias de gestión de la cadena productiva centradas en el caficultor como actor objeto de análisis; se establecen estrategias orientadas a la cosecha y postcosecha del cultivo, como a la gestión de la asociatividad en la región al mediano plazo; a su vez, los mapas de calor estadísticos muestran zonas cultivables de café y de renovación.

**Palabras clave:** estrategias; cultivo de café; analítica de datos

## **Abstract**

*This research proposes management strategies for the coffee production chain using data analytics to improve decision-making in the management of the crop and the producer (coffee grower) as the actor of analysis; the statistical base was formed with data on production, quality, cultivated areas and number of producers of the bean in the department of Norte de Santander; the time window of the data was established from 2013-2024, with which an exploratory analysis was constructed to understand the behavior of the coffee sector in the department and the impacts on productivity and competitiveness at the national level.*

*With the above, the impacts that affect productivity such as geographic dispersion, planting conditions, harvest and post-harvest handling were established; with an accumulated loss of 12% by 2024; which allowed making relational crosses between different associative denominations of producers, cultivated area, available area and crop renewal or transfer; from the above, and to support the strategies, clustering techniques and advanced prediction models were used, segmenting the crops by productive conditions, cultivable area and crop operation.*

*Finally, the study shows the importance of the coffee-growing sector for the region's economy, proposing strategies for managing the production chain focused on the coffee grower as the actor under analysis; strategies oriented to the harvest and post-harvest of the crop are established, as well as to the management of associativity in the region in the medium term; at the same time, the statistical heat maps show coffee cultivation and renewal zones.*

**Keywords:** *strategies; coffee cultivation; data analytics*

## **1. Introducción**

Colombia como productor de café, se posiciona con un alto nivel de exportaciones del grano a diferentes países (Comité de Cafeteros de Norte de Santander, 2022), el sector ha presentado cambios en lo económico, social y ambiental; interviniendo en la estructura, producción y competitividad en los departamentos (Samoggia, et al., 2023); por otro lado, la obsolescencia tecnológica y la baja gestión gubernamental para apoyar el sector afectan la productividad, especialmente en el los pequeños y medianos cultivadores (80% del sector), lo que deja entrever la importancia de desarrollar estrategias de mejoramiento focales (Jacobi, et al., 2024).

La investigación tuvo como objetivo proponer estrategias de gestión para la cadena productiva del café en Norte de Santander, empleando analítica de datos, se utilizaron datos del periodo 2013–2024, que incluyen variables relacionadas con producción, calidad, área cultivada, susceptibilidad del cultivo, procesos de renovación, cumplimiento contractual y defectos en el grano. Mediante el uso de técnicas como análisis exploratorio, conglomerados (clustering) y modelos predictivos (Regresión Lineal Múltiple y Random Forest); se logró caracterizar a los municipios cafeteros según su desempeño y vulnerabilidades; se tomaron como referencia 36 municipios productores de café de la región, para finalmente proponer estrategias diferenciadas por clúster productivo y actor, orientadas a mejorar la productividad, mitigar riesgos

contractuales y fortalecer la sostenibilidad del sistema caficultor regional.

## 2. Marco teórico

### 2.1. Cadena productiva del Café: etapas y actores

Establecer una ruta para una adecuada gestión logística mediante la cadena productiva del café, reduce el desperdicio de subproductos y promueve prácticas sostenibles, estandarizando los procesos, explorando nuevas técnicas para garantizar la consistencia en la calidad del café (Aloo, et al., 2024), asimismo, un desarrollo agrícola acorde a la evolución del mercado (Hidalgo, et al., 2023); enfatizándose en la necesidad de colaboración entre diversos actores, como investigadores, la industria, responsables políticos y comunidades locales (Tamilselvan, et al., 2024); por lo que, la tendencia es su origen único junto con un crecimiento en el consumo internacional, el uso de tecnología y preparación del café (Freitas, et al., 2024).

### 2.2. Factores que afectan la productividad cafetera

Los factores que afectan la producción del café, se deben a aspectos relacionados con el proceso de fermentación, el clima, procedimiento de levaduras/microorganismos, y la calidad del producto final (Bagnulo, et al., 2024). En ese sentido, intervenir en las etapas previas al tostado, como la fermentación, para preservar los atributos sensoriales del café (Barrios, et al., 2023). Asimismo, al considerar el déficit hídrico se proyecta que esta limitación aumente en intensidad y severidad durante escenarios climáticos extremos (León, et al., 2022). Según (Mérillo, et al., 2020) considera una serie de variables interrelacionados para garantizar una producción sostenible y de alta calidad en la industria del café. A continuación, en el cuadro 1 se indican las variables que contienen la relación a tener presente en la producción del café.

Variable	Relación
Condiciones climáticas	Las variaciones en la temperatura, la humedad, la precipitación y la altitud pueden tener un impacto significativo en el crecimiento y desarrollo de las plantas de café
Suelo	La calidad del suelo, incluyendo su composición química, estructura y capacidad de retención de agua, es crucial para el crecimiento saludable de las plantas de café
Plagas y enfermedades	La presencia de plagas como la broca del café y enfermedades como la roya puede causar daños importantes a los cultivos de café y reducir la producción
Prácticas agronómicas	El manejo adecuado de la plantación, incluyendo la poda, la fertilización y el control de malezas, es fundamental para optimizar la producción de café

Cuadro 1. Variables de entrada en el proceso productivo del café

### 2.3. Analítica de datos aplicada a la agricultura

La integración de analítica con sistemas de información geográfica (SIG), sensores remotos e inteligencia artificial ha impulsado el desarrollo de la agricultura de precisión, permitiendo intervenciones personalizadas a nivel de finca, lote o parcela (Wolfert, et al., 2017). En este sentido, la implementación de analítica en la caficultura se convierte en una herramienta para incrementar la rentabilidad, sostenibilidad y competitividad.

### 3. Metodología

La investigación se desarrolló de manera cuantitativa de tipo descriptivo y propositivo, mediante tres fases: análisis exploratorio, analítica de datos y formulación de estrategias, mediante el uso de Python ®©, para el análisis de la cadena productiva del café en el departamento de Norte de Santander, tomando como periodo de estudio los años 2013 a 2024. Se recopiló y sistematizó información relacionada con variables productivas como área cultivada, tasa de renovación, cultivos susceptibles, calidad del grano, defectos identificados, volúmenes de compra y niveles de incumplimiento contractual, proveniente de fuentes institucionales del sector cafetero; las técnicas aplicadas fueron: análisis exploratorio, conglomerados (clustering) y modelos predictivos (Regresión Lineal Múltiple y Random Forest).

### 4. Resultados

#### 4.1. Análisis exploratorio

Para el marco del análisis exploratorio de la cadena productiva del café en el departamento de Norte de Santander, se construyó una base de datos que consolida métricas para la toma de decisiones (Municipio, área en café, cultivos resistentes, cultivos susceptibles, cultivos a renovar, cultivos renovados, tradicional, total cultivos, porcentaje susceptibles y tasa renovación), considerando los municipios con tradición cafetera durante el periodo 2013–2024; en la figura 1 se muestra una concentración del área cultivada en municipios como Toledo y Arboledas, por otro lado, existen municipios con limitaciones climáticas o logísticas que merman el área cultivada del grano.

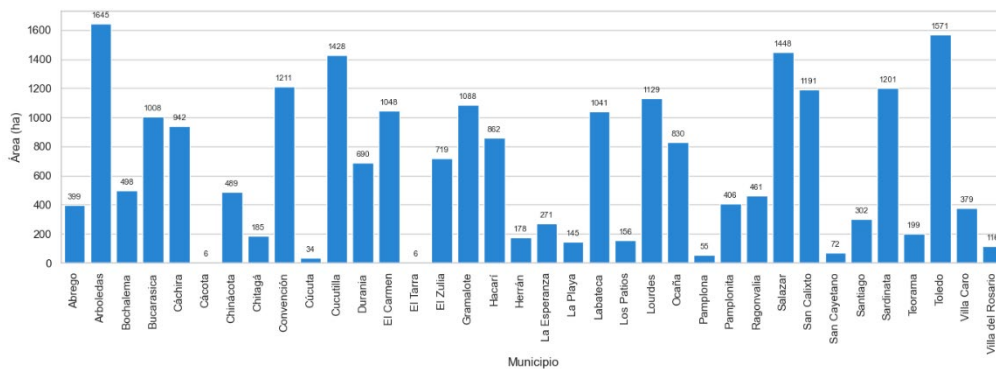


Figura 1. Área total en café por municipio

En términos de cultivos susceptibles a enfermedades o plagas se identifica una heterogeneidad entre los municipios de Teorama, Convención y El Tarra, presentando valores más altos de cultivos, superando el 60% del área cafetera total lo cual representa un riesgo para la sostenibilidad del sector; en tal sentido, se requieren políticas focalizadas en renovación tecnológica, acompañamiento técnico y gestión del riesgo; en contraste, Mutiscua, Pamplonita y Lourdes reportaron proporciones menores al 25%, lo cual indica una mayor estabilidad o esfuerzo en renovación.

Respecto a la tasa de renovación de cultivos municipios como Chitagá, Mutiscua y Pamplona

han liderado procesos de renovación, con tasas superiores al 50% demostrando el uso de prácticas agronómicas más orientadas a la sostenibilidad y adaptación del cultivo, por otro lado, Tibú, Sardinata y Hacarí reportaron tasas inferiores al 20%, asociado a limitaciones estructurales, técnicas o económicas; en una mejora perspectiva de productividad futura se es necesario estrategias de financiamiento, asistencia técnica y estímulos directos al caficultor.

El mapa de calor (figura 2) compara la cantidad de cultivos susceptibles y renovados por municipio, se destaca municipios como Arboledas y Salazar con mayores hectáreas de cultivos de transferencia.

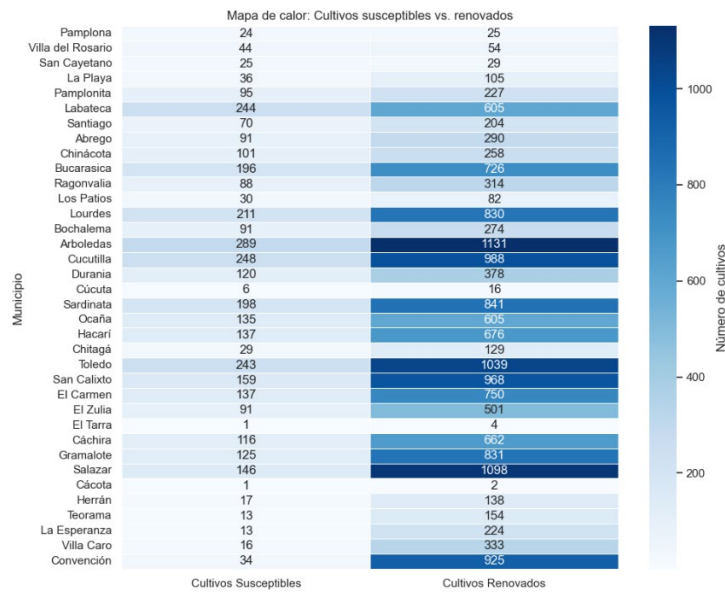


Figura 2. Mapa de calor: cultivos susceptibles vs renovados

Posteriormente, se integraron datos relacionados con las compras de café pergamino seco realizadas por el ente encargado en distintos municipios del departamento, donde se estudiaron variables como el volumen total adquirido, costo de compra, porcentaje de café separado por defectos en taza y los tipos de defecto identificados; en la figura 3 el cruce entre el volumen comprado, calidad del grano y susceptibilidad del cultivo; municipios como Toledo y Ocaña, con volúmenes altos de compra, presentan defectos recurrentes como fermento y moho que afectan directamente la rentabilidad, a su vez, municipios con menores tasas de renovación tienden a reportar mayores porcentajes de café defectuoso, presentando una correlación negativa entre envejecimiento del cultivo y calidad final.

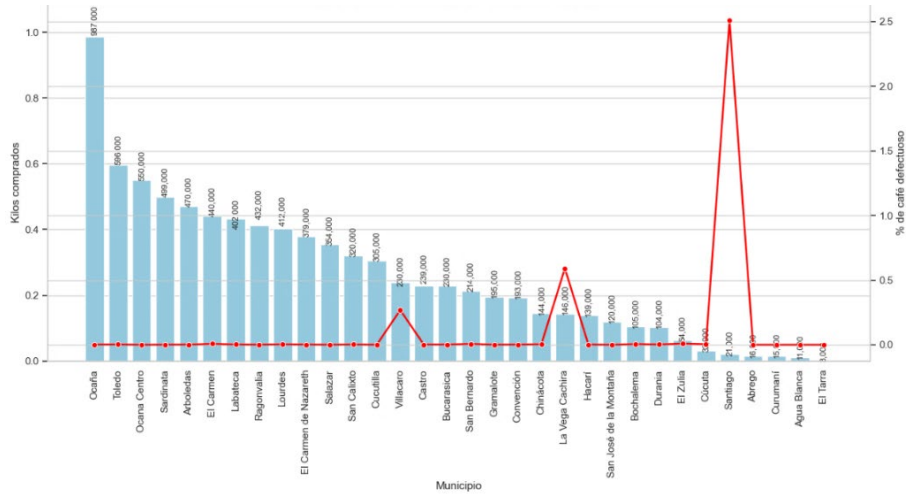


Figura 3. Volumen de café comprado y % defectuoso por municipio

Otro análisis se da en la distribución de kilogramos por tipo de contrato y tipo de caficultor, se presenta un incumplimiento en los contratos "*Practicas Precio Fijo*", donde los caficultores no asociados representaron un incumplimiento mayor; en cuanto a los contratos "*Futuros*", los incumplimientos son menores, pero siguen la misma tendencia. En proporción de incumplimientos por tipo de caficultor, se determina que el 72,9% del total de kg incumplidos proviene de caficultores no asociados, mientras que solo el 27,1% corresponde a asociados, lo que llevaría a determinar estrategias de fortalecimiento organizativo en la región.

Con lo anterior, la pérdida económica se relaciona a un incremento considerable en los años 2021 y 2022; las pérdidas generadas por incumplimientos ascendieron a \$1.113.405.000 en el 2021, mientras que en 2022 fueron de \$1.681.333.000, representando un aumento del 51% lo que afecta directamente la rentabilidad del sector caficultor en tal sentido se deben implementar mecanismos que mejoren el cumplimiento de los caficultores no asociados.

#### 4.2. Segmentación de cultivos (*Clustering*)

Con el fin de identificar patrones homogéneos entre municipios del departamento en función de variables clave como el área cultivada, porcentaje de cultivos susceptibles, tasa de renovación y cantidad absoluta de cultivos renovados; la segmentación de los cultivos cafetaleros se realizó mediante la técnica de agrupamiento no supervisado K-Means, previamente, las variables fueron normalizadas y se aplicó un análisis de componentes principales (PCA) para reducir la dimensionalidad y facilitar la visualización de los clúster, como se presenta en la figura 4.

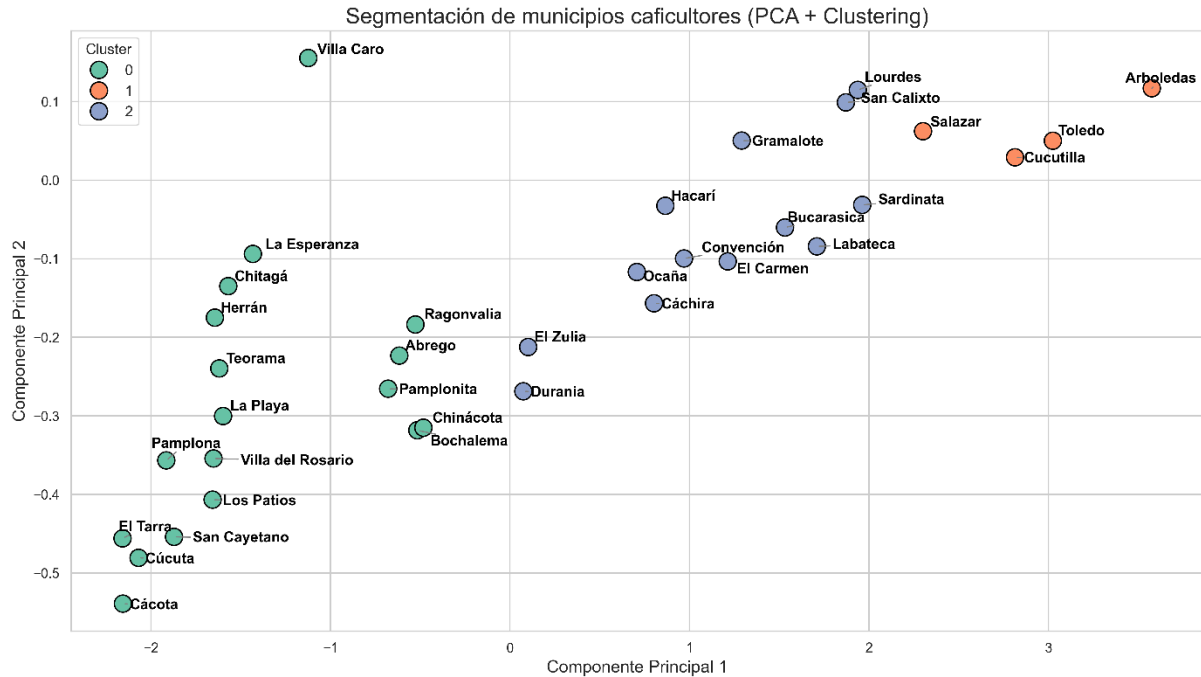


Figura 4. Segmentación de municipios caficultores (PCA + Clustering)

El modelo estableció tres grupos de municipios con características diferenciadas como se detalla en el cuadro 2, sobre un plano bidimensional utilizando los dos primeros componentes principales, los cuales explicaron el 82.4% de la varianza total del sistema, permitiendo una representación visual confiable de los datos.

Clúster	Número de municipios	Características
Productores consolidados	20 municipios	Alta tasa de renovación, área cultivada media, nivel intermedio de cultivos susceptibles
Productores emergentes	18 municipios	Baja tasa de renovación, mayor presencia de cultivos susceptibles, menor área cultivada
Productores vulnerables	7 municipios	Alta concentración de área cultivada, menor porcentaje de cultivos susceptibles, renovación moderada

Cuadro 2. Distribución por clúster

En el cuadro 3, los municipios del clúster 2 presentan condiciones favorables en términos de productividad y renovación, lo cual los ubica como referentes dentro del ecosistema caficultor regional, por el contrario, los municipios del clúster 1 se identifican como territorios con alto nivel de vulnerabilidad productiva, debido a una baja tasa de renovación y una elevada presencia de cultivos susceptibles.

Variable	Productores consolidados	Productores emergentes	Productores vulnerables
Área cultivada (ha)	128.6	75.3	223.4
Cultivos susceptibles (ha)	42.7	61.9	19.6
Cultivos renovados (ha)	21.5	9.8	31.4
Tasa de renovación (%)	27.4	12.9	38.7

Cuadro 3. Promedio por clúster para cada variable

#### 4.3. Modelos de predicción

Con el objetivo de anticipar el comportamiento de los cultivos renovados en función de variables productivas y contextuales, se desarrollaron modelos predictivos utilizando técnicas de regresión lineal múltiple y Random Forest. Los modelos se entrenaron con variables independientes como el área sembrada, número de cultivos susceptibles, tasa de renovación histórica y porcentaje de susceptibilidad; por lo que los resultados obtenidos muestran que el modelo de Random Forest presentó un mejor ajuste predictivo, alcanzando un coeficiente de determinación  $R^2$  más alto y un menor error cuadrático medio (MSE) en comparación con la regresión lineal tradicional, demostrando la capacidad del modelo no lineal para capturar relaciones complejas entre las variables explicativas y la variable objetivo.

Asimismo, se evaluó el desempeño de los dos modelos de predicción: regresión lineal y Random Forest, utilizando dos métricas comunes: el coeficiente de determinación  $R^2$  y el error cuadrático medio (MSE). Como se observa en el cuadro 4, el modelo de regresión lineal presentó un valor de  $R^2$  negativo (-1.033), lo que indica un desempeño deficiente, adicionalmente, su error cuadrático medio (MSE) fue considerablemente alto (205160.09), lo que refleja una baja precisión en las predicciones. En contraste, el modelo de Random Forest mostró un desempeño superior, con un  $R^2$  de (0.973) y un MSE de solo (2758.62), este modelo se adapta de forma efectiva a la relación entre las variables, siendo adecuado para los datos analizados.

Modelo	$R^2$ Score	Error Cuadrático Medio (MSE)
Regresión Lineal Múltiple	-1.033	205160.09
Random Forest	0.973	2758.62

Cuadro 4. Evaluación del desempeño de los modelos de predicción

En cuanto a la comparación del desempeño de ambos modelos, mediante su  $R^2$  y MSE en la figura 5, se visualizan las diferencias en la capacidad de predicción entre los modelos el cual contiene dos ejes verticales: uno para el  $R^2$  (en azul) y otro para el MSE (en verde); la barra azul representa el  $R^2$  de cada modelo, mientras que la línea verde muestra el MSE correspondiente.

Por consiguiente, se sugiere que mediante la predicción de Random Forest se muestra como una técnica adecuada para anticipar comportamientos en el proceso de compras, como el volumen adquirido, la frecuencia o incluso el riesgo de incumplimiento, la implementación de este tipo de modelos predictivos contribuye de forma relevante en la toma de decisiones estratégicas por parte de asociaciones de caficultores, permitiéndoles mejorar su producción, planificar la distribución y ajustar las condiciones comerciales según el perfil de riesgo y comportamiento de los compradores.

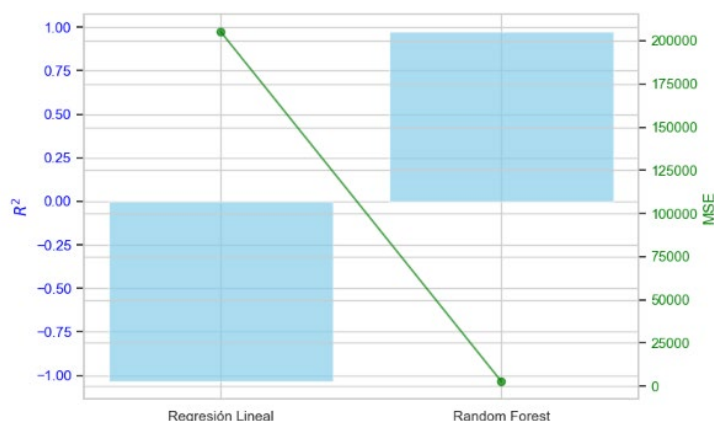


Figura 5. Comparación de modelos: Regresión Lineal vs Random Forest

#### 4.4. Propuesta de estrategias

A partir del análisis de clústeres y los niveles de incumplimiento contractual, se realizó la segmentación para la asignación de recursos y la sostenibilidad del sistema como una herramienta que permite tomar decisiones para los actores de conforman la cadena.

Se formularon estrategias diferenciadas para cada clúster, cruzando las salidas del modelo de agrupamiento (k-means) con variables como el volumen de producción, acceso a asistencia técnica y riesgo de incumplimiento contractual; alineando intervenciones técnicas, contractuales y sociales según el perfil de riesgo y capacidades productivas.

Clúster	Características	Estrategia	Conceptualización
Productores consolidados (alto volumen, bajo riesgo)	<ul style="list-style-type: none"> <li>Producción &gt; 4000 kg/año</li> <li>Alta asistencia técnica</li> <li>Riesgo de incumplimiento &lt; 15%</li> </ul>	<ol style="list-style-type: none"> <li>Programas de fidelización con beneficios</li> <li>Alianzas con compradores internacionales</li> <li>Capacitación en diferenciación del producto</li> <li>Mejores condiciones contractuales por cumplimiento</li> </ol>	<ul style="list-style-type: none"> <li>Representan el 32% de los productores</li> <li>Aportan el 48% del volumen</li> <li>Riesgo de incumplimiento menor al 15%</li> </ul>
Productores emergentes (riesgo medio)	<ul style="list-style-type: none"> <li>Producción entre 2000-3000 kg/año</li> <li>Asistencia técnica ocasional</li> <li>Riesgo de incumplimiento 25-35%</li> </ul>	<ol style="list-style-type: none"> <li>Incentivos para asistencia técnica continua</li> <li>Seguro de cosecha condicionado al cumplimiento</li> <li>Microfinanzas rurales específicas</li> <li>Capacitación contractual</li> </ol>	<ul style="list-style-type: none"> <li>Generan el 45% de los incumplimientos</li> <li>Mejora del cumplimiento en 12% con asistencia técnica</li> <li>Segmento de transición clave</li> </ul>
Productores vulnerables (bajo volumen, alto riesgo)	<ul style="list-style-type: none"> <li>Producción &lt; 1500 kg/año</li> <li>Poco acceso a crédito</li> <li>Riesgo de incumplimiento &gt; 50%</li> </ul>	<ol style="list-style-type: none"> <li>Intervención social integral</li> <li>Contratos adaptativos y flexibles</li> <li>Subsidios para transición agroforestal</li> <li>Prioridad en inversión pública o de ONGs</li> </ol>	<ul style="list-style-type: none"> <li>23% de productores</li> <li>Generan el 62% de los incumplimientos</li> <li>Alta vulnerabilidad social y económica</li> </ul>

Cuadro 5. Estrategias según los resultados de los modelos de predicción

## 5. Conclusiones

Durante el periodo 2013–2024, la productividad cafetera en Norte de Santander presentó una disminución acumulada del 12%, demostrando una pérdida sostenida de rendimiento por factores como el envejecimiento de cultivos y baja tasa de renovación; además, se identificó una pérdida del 20% en la presencia de cultivos activos en la región en los últimos 11 años, afectando la competitividad del sector, especialmente en municipios con alta proporción de cultivos susceptibles como Teorama y Convención, entre los cuales superan el 60% del total.

Desde la gestión organizativa, los caficultores no asociados concentraron el 72,9% de los incumplimientos contractuales, presentando una pérdida económica estimada en \$1.681 millones para 2022, lo que representa un aumento del 51% frente a 2021 por lo que se debe fortalecer las figuras asociativas.

Para el planteamiento de las estrategias se emplearon técnicas de clustering lo que permitió segmentar los productores en tres grupos, donde, el clúster consolidados generaron el 48% del volumen total de café con un riesgo de incumplimiento menor al 15%; mientras que el clúster vulnerables concentró el 62% de los incumplimientos, es así que el modelo predictivo Random Forest logro determinar escenarios productivos y riesgos contractuales con un coeficiente de determinación  $R^2$  de 0.973 y un MSE de solo 2758.62, sustentando las estrategias propuestas.

## 6. Referencias

- Aloo, F., Mutuku, A., & Nyang'au, S. (2024). Sustainability-oriented coffee value chains: Role of quality assurance and collaboration. *Sustainability*, 16(4), 2651. <https://doi.org/10.3390/su16042651>
- Bagnulo, M., Rivera, E., & Palacios, J. (2024). Fermentation and quality management in specialty coffee production. *Food Chemistry*, 421, 136120. <https://doi.org/10.1016/j.foodchem.2023.136120>
- Barrios, H., Mendoza, C., & Rojas, D. (2023). Yeasts and microorganisms in pre-roasting coffee processes: a sensory preservation perspective. *Journal of Food Microbiology*, 102, 104330. <https://doi.org/10.1016/j.fm.2023.104330>
- Comité de Cafeteros de Norte de Santander. (2022). Generalidades de la caficultura en el departamento Norte de Santander. Federación Nacional de Cafeteros. <https://nortedesantander.federaciondecafeteros.org/cafe-de-norte-de-santander/>
- Freitas, C., Souza, R., & Tavares, G. (2024). Emerging trends in single-origin coffee: consumption, preparation, and technology. *Journal of Coffee Research*, 18(2), 95-110.
- Hidalgo, C., Torres, M., & Medina, D. (2023). Agricultural development strategies aligned with market evolution in Latin America. *Agroecología y Desarrollo*, 19(3), 85–96.
- Jacobi, J., de Castelberg, S., Yeretziyan, Ch., et al. (2024). Making specialty coffee and coffee-cherry value chains work for family farmers' livelihoods: A participatory action research approach. *World Development Perspectives*, 33, 100551. <https://doi.org/10.1016/j.wdp.2023.100551>
- León, J., Unigarro, A., & Balaguera, A. (2022). El impacto del déficit hídrico en cultivos de café: un análisis de escenarios extremos. *Revista Colombiana de Climatología*, 14(1), 44-58.
- Mérillon, J. M., & Ramawat, K. G. (2020). *Bioactive molecules in food*. Springer, Cham. <https://doi.org/10.1007/978-3-319-78030-6>



- Samoggia, A., & Fantini, A. (2023). Revealing the governance dynamics of the coffee chain in Colombia: A state-of-the-art review. *Sustainability*, 15(18), 13646. <https://doi.org/10.3390/su151813646>
- Tamilselvan, R., Kumar, V., & Lakshmi, R. (2024). The role of multi-stakeholder collaboration in sustainable agricultural chains. *Agricultural Systems*, 212, 103723. <https://doi.org/10.1016/j.agsy.2023.103723>
- Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big data in smart farming – A review. *Agricultural Systems*, 153, 69–80. <https://doi.org/10.1016/j.agsy.2017.01.023>

## Sobre los autores

- **Karen Julieth Castellanos Rodríguez:** Ingeniera Industrial, Magister en Ingeniería Industrial. [karen.castellanos@unipamplona.edu.co](mailto:karen.castellanos@unipamplona.edu.co)
- **Óscar Mayorga Torres:** Ingeniero Industrial, Magister en Ingeniería Industrial, Profesor Universidad Francisco de Paula Santander. [oscarmtorres@ufps.edu.co](mailto:oscarmtorres@ufps.edu.co)

---

Los puntos de vista expresados en este artículo no reflejan necesariamente la opinión de la Asociación Colombiana de Facultades de Ingeniería.

Copyright © 2025 Asociación Colombiana de Facultades de Ingeniería (ACOFI)