

MUTIS



Chachafruto
Erythrina edulis
Benjamín Cárdenas Villavejana, 2018

Universidad de Bogotá Jorge Tadeo Lozano
Jefatura de Publicaciones
Carrera 4 n.º 23-76, Módulo 29, of. 203
Piso 2, Bogotá, D. C., Colombia
Tel: (571) 242 7030, Exts. 3120-3132

Rector
CARLOS SÁNCHEZ GAITÁN

Vicerrector Académico
ANDRÉS FRANCO HERRERA

Vicerrectora Administrativa
LILIANA ÁLVAREZ REVELO

Director de Investigación, Creación y Extensión
LEONARDO PINEDA SERNA

Jefe de Publicaciones
MARCO GIRALDO BARRETO

Coordinador Editorial
JUAN CARLOS GARCÍA SAENZ

Coordinación gráfica y diseño y diagramación
LUIS CARLOS CELIS CALDERÓN

Corrector de estilo
LEONARDO PAIPILLA PARDO

Coordinación editorial y revisión
SYLVANA SILVANA BLANCO ESTRADA

Distribución y ventas
SANDRA GUZMÁN

Asistente administrativa
MARÍA TERESA MURCIA

Mutis es una publicación electrónica semestral de ciencia e investigación editada por la Facultad de Ciencias Naturales e Ingeniería de la Universidad de Bogotá Jorge Tadeo Lozano. Los artículos publicados son responsabilidad de sus autores y no comprometen la posición editorial de *Mutis*.

Editor

GERARD OLIVAR TOST

Editor de sección

JAVIER HERNÁNDEZ FERNÁNDEZ

Comité Editorial

MICHAEL AHRENS

Universidad de Bogotá Jorge Tadeo Lozano, Colombia

JUAN PABLO APARICIO

Universidad Nacional de Salta, Argentina

JUDITH A. CHERNI

Imperial College London, Reino Unido

PEDRO GAJARDO

Universidad Técnica Federico Santa María, Chile

LEONARDO MARIÑO-RAMÍREZ

NCBI - National Center for Biotechnology Information,
Estados Unidos

LUIS RABELO

Universidad Central de la Florida, Estados Unidos

LORENA NOVOA APONTE

National Institutes of Health, Estados Unidos

MYRON SPECTOR

Harvard University, Estados Unidos

ENRIC TRULLOLS FARRENY

Universidad Politécnica de Cataluña, España

Comité Científico

ALBA GRACIELA ÁVILA

Universidad de los Andes, Colombia

SILVIO ALEJANDRO LÓPEZ PASOS

Universidad Colegio Mayor de Cundinamarca,
Colombia

MÓNICA PUYANA HEGEDUS

Universidad de Bogotá Jorge Tadeo Lozano, Colombia

Actualmente el contenido está indexado e incluido en Google Scholar, Academia.edu, BASE, Dialnet, Dimensions, Directory of Research Journals Indexing, Electronic Journals Library, Latindex MIAR, Mendeley, PERIODICA, Publindex - Categoría C, PKP index, ROAD, Ulrich's Periodicals Directory, AmelICA, DOAJ, Ebsco (fuente académica), Redib



CONTENIDO

Vol. 11 n.º 1 – I. Edición 2021

EDITORIAL

Biología sintética y SARS-Cov-2

Javier Hernández Fernández

5

ARTÍCULOS DE INVESTIGACIÓN – RESEARCH ARTICLES

Efecto de los elementos menores en la calidad molinera del arroz (*Oryza sativa L.*) variedad F-2000

Effect of Minor Elements on the Milling Quality of F-2000 Rice Variety (*Oryza sativa L.*)

José Alirio Hernández Quiñónez, Seir Antonio Salazar Mercado y Edgar Alfonso Rodríguez Araújo

8

Análisis multicriterio de plantas medicinales cultivadas en la vereda Ceylan del municipio de Viotá, Cundinamarca, Colombia

Multi-criteria Analysis of the Medicinal Plants Grown in Ceylan, municipality of Viotá, Cundinamarca, Colombia

Yuri Alicia Chávez-Plazas, María Lucero Ramírez-Mahecha y Judith Elena Camacho-Kurmen

22

¿Ganadería o conservación? Caracterización del conflicto ambiental en la Reserva Forestal Protectora “El Malmo” (Tunja, Boyacá, Colombia)

Livestock or Conservation? Characterization of the Environmental Conflict in the Protective Forest Reserve “El Malmo” (Tunja, Boyacá, Colombia)

Albaluz Ramos Franco, Inti Alba Vargas, Leyde Cabrera Ruiz y Camila Diaz Rojas

37

DOSSIER. ANALÍTICA DE DATOS

Detección de contratistas multiobjeto mediante minería de textos para focalizar el ejercicio del control y vigilancia fiscal

Detection of Multi-object Contractors through Text Mining to Targeting the Exercise of Fiscal Control and Surveillance

Manuel Francisco Dulce Vanegas y Adam Beltrán Gómez

48

Predicción de la serie temporal del indicador bancario de referencia (IBR) con redes neuronales

Predicting the Benchmark Banking Indicator (IBR) Time Series through Neural Networks

Germán Enrique Coy Mondragón, Óscar Granados y Olmer Garcia-Bedoya

65

Análítica de grafos para identificar entidades relevantes y comunidades en Mercado Libre: un estudio de caso

Use of Graph Analytics to Identify Relevant Entities and Communities in Mercado Libre: A Case Study

Gerson Enrique Carrillo Gelvez y Ixent Galpin

77

Instrucciones para autores

96

EDITORIAL

Vol. 11 (1) pp. 5-7, enero-junio del 2021

Biología sintética y SARS-cov-2

La biología sintética, desde hace ya varios años —tal vez una década—, ha generado la idea de la creación de la vida desde cero. Los investigadores han diseñado experimentos para sintetizar proteínas en un papel de filtro. Se trasplantó el genoma bacteriano entre diferentes especies (*Mycoplasma mycoides* a *Mycoplasma caplicolium*), se creó un genoma sintético y se puso a funcionar en una bacteria (*Mycoplasma laboratorium*), cuyo material genético había sido previamente removido. Además, actualmente se están desarrollando microorganismos con un genoma mínimo, a los que podrían añadirse genes y de esta manera realizar funciones específicas, tales como la biodegradación de agentes tóxicos del ambiente, la reducción de CO₂ en cantidades que permitan minimizar el calentamiento global y la producción de bioetanol o biodiésel (Hernández-Fernández, 2012).

La biología sintética es un nuevo campo de la biología y la ingeniería que utiliza metodologías de estas y de otras disciplinas, y que se define como el diseño y la creación de estructuras y sistemas biológicos no existentes en la naturaleza. Este campo produce nuevas rutas metabólicas, construye células mínimas, crea sistemas biológicos con estructuras bioquímicas alternativas, modifica la maquinaria celular y elabora dispositivos biológicos. Su objetivo es la creación de nuevos artefactos que respondan a estímulos de forma programada, controlada y confiable. Esta disciplina es ahora una herramienta esencial para una amplia gama de aplicaciones (ETC Group, 2007).

En el descubrimiento de nuevas drogas, la biología sintética está mostrando diferencias apreciables para activar el desarrollo de fármacos y reducir costos. Una de sus áreas de aplicación es la secuenciación de genomas, proceso esencial para la investigación farmacéutica. Hoy día, esta área trabaja con más información genómica que nunca. Así, el inusitado aumento de datos de secuencia permite diseñar medicamentos clínicamente más eficaces (Maclean, 2021).

La biología sintética ha focalizado a los científicos que trabajan en farmacéutica en el área de síntesis de fragmentos de ADN, ARN, clones o genotecas de variantes completas. Los investigadores utilizan plataformas automatizadas para la síntesis rápida de ácidos nucleicos (ARN o ADN), y con ello producir vacunas candidatas contra virus, como el SARS-COV-2, mucho más rápido de lo que generalmente es posible (Maclean, 2021).

¿Para qué sirve y cómo funciona la biología sintética?

Esta disciplina ha puesto al servicio de la ciencia y la farmacología sistemas libres de células que permiten evaluar maquinaria biológica, evitando variables que pueden llevar a confusiones cuando se utiliza un sistema celular vivo. Entonces, se han utilizado

membranas lipídicas construidas en laboratorio que imitan a una célula, sin serlo. Estos acercamientos han abierto la puerta a nuevas ideas para manipular sistemas con el propósito de descubrir nuevas rutas que permitan la producción de medicamentos y vacunas importantes (Macleay, 2021).

Avances recientes en el campo han dado la posibilidad de sintetizar ADN y ARN con una alta precisión y pureza increíble. Este hecho ha permitido realizar experimentos que, a la postre, nos han llevado al descubrimiento de las vacunas recombinantes de ADN y ARN. Los ácidos nucleicos sintetizados se pueden usar en ingeniería de proteínas y en producción de anticuerpos. Estas aplicaciones son muy importantes en todos los descubrimientos realizados recientemente. También se han ensamblado secuencias de nanocuerpos que han resultado exitosos, así como los candidatos a vacunas (SGI-DNA, 2020).

La unión entre la biología sintética y el desarrollo de vacunas ha dado como resultado múltiples candidatos a vacunas contra el virus SARS-cov-2, de lo que ya todos conocemos. Millones de personas en todo el mundo han sido vacunadas, obteniendo protección o minimizando el cuadro clínico de la infección versus aquellas que se han negado a la inmunización por diferentes motivos.

La biología sintética estuvo disponible durante otros brotes epidémicos, como la gripe aviar H7N9, que nunca se convirtió en pandemia. Así mismo, los sistemas de automatización para construir DNA y RNA se produjeron al tiempo que apareció en Wuhan, China, el virus SARS-cov-2, un nuevo coronavirus que causa la COVID-19. En unas pocas semanas las candidatas a vacunas ya se habían evaluado en ratones (SGI-DNA, 2020).

La posibilidad de producir vacunas que cubran a todas las variantes con mutaciones puntuales —como la variante delta— es totalmente posible, si es que las vacunas disponibles ya no lo están haciendo.

Lo más posible es que el virus continúe variando y nuevas vacunas sigan produciéndose para poder inmunizar a la población. Ya se habla de una tercera dosis, y si el virus no cambia para volverse inofensivo, se hablará de cuarta dosis y otras adicionales. Todo esto se podrá ver a velocidades mayores de las ya vistas, y es casi seguro que la biología sintética encontrará la ruta para aniquilar este virus.

Javier Hernández Fernández

Editor Ciencias Biológicas – Revista Mutis
<https://orcid.org/0000-0001-8442-9266>
<https://doi.org/10.21789/22561498.1787>

REFERENCIAS

- ETC Group. (2007, enero 16). *Extreme genetic engineering, an introduction to synthetic biology*. ETC Group. <http://www.etcgroup.org/node/602>
- Hernández-Fernández, J. (2012). Biología sintética ¿creando la vida desde cero? *Expediitio*, 11, 67-81. <https://revistas.utadeo.edu.co/index.php/EXP/article/view/753>
- Maclean, K. (2021, mayo 27). *Acelerando el descubrimiento de fármacos a través de la biología sintética*. Tecnología Redes. <http://www.news-courier.com/drug-discovery/articles/accelerating-drug-discovery-through-synthetic-biology-335325>
- SIGI-DNA. (2020, febrero 18). *SIGI-DNA equipping researchers with critical tools to combat coronavirus disease (COVID-2019)*. SIGI-DNA.

Efecto de los elementos menores en la calidad molinera del arroz (*Oryza sativa* L.) variedad F-2000

Effect of Minor Elements on the Milling Quality of F-2000 Rice Variety (*Oryza sativa* L.)

José Alirio Hernández Quiñónez^{ac}, Seir Antonio Salazar Mercado^{bd}
y Edgar Alfonso Rodríguez Araújo^{ae}

^a Departamento de Agronomía, Facultad de Ciencias Agrarias y del Ambiente, Universidad Francisco de Paula Santander, Cúcuta, Colombia

^b Departamento de Biología, Facultad de Ciencias Básicas, Universidad Francisco de Paula Santander, Cúcuta, Colombia

^c <https://orcid.org/0000-0002-5177-962X>

^d <http://orcid.org/0000-0002-3287-703X>

^e edgarrodriguez@ufps.edu.co | <https://orcid.org/0000-0003-3868-1705>

RESUMEN

Esta investigación evaluó la aplicación de fertilizantes que contenían elementos menores como mejoradores de la calidad molinera del arroz (*Oryza sativa* L.) en un suelo clasificado como Typic Xerofluvents. Se valoró un área de 2,5 ha, divididas en cuatro piscinas cultivadas con arroz variedad Fedearroz 2000. Se utilizó un diseño de bloques completos al azar con cinco tratamientos y cuatro repeticiones. Las variables de producción evaluadas fueron: densidad de plantas ($46,27 \pm 320$ plantas/m²), macollamiento ($59,48 \pm 821,26$ macollas/m²), número panículas/área ($8,53 \pm 440$ panículas/m²), número espiguillas/panícula ($2,76 \pm 12,3$ espiguillas/panícula), número de granos totales ($15,44 \pm 138,77$ granos), número de granos llenos ($7,88 \pm 116$ granos) y número de granos vanos ($19,88 \pm 22,28$ granos).

Para evaluar el componente de molinearía se utilizaron 500 gramos de granos y se valoraron las siguientes variables: porcentaje de vaneamiento ($15,24 \pm 18,48$), masa blanca ($70,57 \pm 70,68$), grano entero ($446,9 \pm 59,65$), cristal ($176,93 \pm 25,24$ %), harina ($79 \pm 12,82$), cáscara ($7,89 \pm 20,85$) y grano yesado ($82,08 \pm 2,08$). Para determinar el rendimiento ($37,19 \pm 7,689$ Kg/ha) se cosechó y se pesó cada tratamiento por separado. Se presentaron diferencias significativas ($p < 0,0001$) entre los tratamientos, siendo T2 (fertilización tradicional de NPK + fertilización edáfica con elementos menores) el que presentó mejor comportamiento entre variables de producción, mientras que

Citation: Hernández-Quiñónez, J. A., Salazar-Mercado, S. A. y Rodríguez-Araújo, E. A. (2021). Efecto de los elementos menores en la calidad molinera del arroz (*ORYZA SATIVA* L.) variedad F-2000. *Mutis*, 11(1), 8-21. doi: <https://doi.org/10.21789/22561498.1711>

Recibido: Diciembre 7, 2020. **Aceptado:** Diciembre 30, 2020.

Copyright: ©2021. Hernández-Quiñónez, J. A., Salazar-Mercado, S. A. y Rodríguez-Araújo, E. A. (2021). This is an open-access article, which permits unrestricted use, distributions and reproduction in any medium, provided the original author and source are credited.

Competing Interests: The authors have no conflict of interest.



T5 (fertilización tradicional de NPK + fertilización foliar con elementos menores + fertilización foliar con silicio) reportó las mejores variables de molinería. La fertilización con elementos menores contribuye a mejorar la calidad molinera y el rendimiento del arroz, puesto que fortalece el sistema inmune de la planta, combate agentes bióticos y evita los granos vacíos.

Palabras clave: fertilización edáfica, fertilización foliar, macolla, panícula, grano entero.

ABSTRACT

The research evaluated the application of fertilizers containing minor elements as milling quality improvers of rice (*Oryza sativa* L) in Typic Xerofluvents soils. We assessed an area of 2.5 ha divided into four ponds and cultivated with Fedearroz 2000 rice variety. A randomized complete block design with five treatments and four replications was used. The production variables evaluated were: plant density (46.27 ± 320 plants/m²), tillering (59.48 ± 821.26 tillers/m²), number of panicles/area (8.53 ± 440 panicles/m²), number of spikelets/panicle (2.76 ± 12.3 spikelets/panicle), number of total grains (15.44 ± 138.77 grains), number of filled grains (7.88 ± 116 grains), and number of vain grains (19.88 ± 22.28 grains). 500 grams of kernels were used to evaluate the milling component, assessing the following variables: percentage of spike sterility (15.24 ± 18.48), white mass (70.57 ± 70.68), whole grain (446.9 ± 59.65), crystal ($176.93 \pm 25.24\%$), flour (79 ± 12.82), husk (7.89 ± 20.85), and plaster grain (82.08 ± 2.08). Each treatment was harvested and weighed separately in order to determine yield ($37.19 \pm 7,689$ kg/ha). There were significant differences ($p < 0.0001$) among the treatments, with T2 (traditional NPK fertilization + soil fertilization with minor elements) showing the best performance among production variables and T5 among milling variables (traditional NPK fertilization + foliar fertilization with minor elements + foliar fertilization with silicon). Fertilization with minor elements improves milling quality and rice yield by strengthening the plant's immune system, thus combating biotic agents and avoiding empty grains.

Keywords: Soil fertilization, foliar fertilization, bunch, panicle, whole grain.

INTRODUCCIÓN

El arroz es una planta liliopsida que pertenece a la familia de las gramíneas. Existen 19 especies, siendo el arroz común (*Oryza sativa* L.) la más importante, puesto que es un alimento básico de la canasta familiar para más de la mitad de la población mundial y el segundo cereal con mayor área cosechada. Su producción promedio anual alcanza, aproximadamente, 650 millones de toneladas (Barona, 2010; Hernández-Rodríguez *et al.*, 2014; Kato & Katsura, 2014; Rives *et al.*, 2007). Más del 85 % del área mundial plantada con arroz se riega, principalmente por inundaciones, porque este método aumenta los rendimientos (Botta *et al.*, 2015).

La calidad molinera se define a partir de la cantidad de grano que permanece entero o en tres cuartos de su tamaño después de ser sometido al proceso de descascarado y pulido. La calificación del arroz molinado

en términos de rendimiento en molino y calidad no depende únicamente del tipo o condición del equipo o de la destreza del operario, sino que está determinada en gran parte por la calidad del arroz en cáscara (Gnanamanickam, 2009; Mohanty, 2013; Muthayya *et al.*, 2014). El desempeño de este producto durante la molienda está influenciado en gran medida por las características del lote de arroz, incluyendo factores genéticos y ambientales, condiciones de procesamiento y el tipo de molino utilizado (FAO, 2009, 2014).

Entre los factores que influyen en la calidad de molienda se encuentra la disponibilidad de nutrientes en el suelo, particularidad que define un buen lote de arroz. Sin embargo, la mayoría de los suelos presentan deficiencias de ciertos elementos o en ocasiones estos no se encuentran disponibles, lo que hace necesaria la aplicación de fertilizantes para suplir

requerimientos de tipo nutricional (Dutta *et al.*, 2016; Fairhurst & Witt, 2002; Serraj *et al.*, 2011).

El grupo de nutrientes se encuentra conformado por nitrógeno (N), fósforo (P) y potasio (K), denominados macro nutrientes primarios, seguidos de azufre (S), calcio (Ca) y magnesio (Mg), nombrados como macro nutrientes secundarios (Alshaal & El-Ramady, 2017; Cakmak & Kutman, 2017; IFPRI, 2014). Lo anterior se ha determinado según la cantidad de cada nutriente utilizada por la planta y la frecuencia con que en la práctica es necesaria su aportación al cultivo. No obstante, existen elementos que son absorbidos en mínimas cantidades, pero que desempeñan funciones fisiológicas muy importantes. Estos son los elementos menores hierro (Fe), cobre (Cu), zinc (Zn), manganeso (Mn), molibdeno (Mo), boro (B) y silicio (Si), en el caso particular del arroz (Bashir *et al.*, 2013; Das *et al.*, 2013; Virgili, 1996; Zhang *et al.*, 2017).

Se ha comprobado que la aplicación de elementos menores incrementa la producción de arroz en porcentajes significativos debido a que estos forman parte de enzimas que ayudan a la asimilación de los demás elementos, lo cual se traduce en un mayor aprovechamiento de los nutrientes por parte de la planta. Sin embargo, los estudios realizados en campo solo se limitan a la parte productiva (Alamdari & Mobasser, 2014; Álvarez *et al.*, 2008; Bouis & Saltzman, 2017; Meena & Fathima, 2017; Phattarakul *et al.*, 2012; Shaygany *et al.*, 2012; Siddika *et al.*, 2016; Yuan *et al.*, 2012). Por tal motivo, el presente estudio analiza cuatro tratamientos con elementos menores y su influencia en la calidad molinera del arroz.

MATERIALES Y MÉTODOS

El experimento se realizó en la vereda Borriqueros, ubicada a 7°56' de latitud norte y 72°36' de longitud oeste, en el municipio del Zulia, departamento de Norte de Santander, Colombia. La zona presenta un clima semiárido cálido (según clasificación climática de Köppen: BSh), con altitud sobre el nivel del mar de 220 metros, temperatura media anual de 28 °C y humedad relativa de 58 %. La investigación se realizó en un área de 2,5 ha, las cuales fueron divididas en cuatro piscinas

cultivadas con arroz variedad Fedearroz 2000, estableciendo cinco bloques de 36 m² para cada tratamiento.

Preparación del lote para la siembra

Se realizó la nivelación del lote mediante la utilización de pala láser. Antes de la siembra se implementó un pase de rastra para mejorar la aireación del suelo e incorporar los residuos del control de malezas. La prueba de germinación se realizó con 100 semillas tomadas al azar con tres replicas (Mercado *et al.*, 2020; Salazar *et al.*, 2020a; Salazar *et al.*, 2020b; Salazar-Mercado *et al.*, 2020). Todas estas semillas fueron previamente sumergidas en agua durante dos días para hidratarlas e inducir la germinación. La siembra se ejecutó con una densidad de 160 kg/ha de forma manual al voleo.

Plan de fertilización

La fertilidad del lote se basó en los resultados arrojados por el plan integral de nutrición con tecnología satelital, conocido como PINSAT. En la primera fertilización (tabla 1) se aplicó fósforo y nitrógeno y se adicionaron elementos menores y una fracción de elementos secundarios. Así mismo, en el tratamiento T3 se realizó la primera aplicación del silicio edáfico. En la segunda fertilización se realizó la segunda aplicación de silicio edáfico en el T3. Igualmente, se realizó la aplicación de fósforo y nitrógeno como elementos esenciales, más la adición de magnesio y azufre. Para la tercera fertilización se adicionó la tercera fracción de nitrógeno y potasio más la última fracción de calcio y boro. En la cuarta fertilización se adicionó nitrógeno y potasio en partes iguales, con pequeñas fracciones de azufre y magnesio. Las aplicaciones foliares (tabla 2) se realizaron 45 y 55 días después del germinado (d. d. g.) en máximo macollamiento y tras la aparición del anillo verde del arroz.

El ensayo se realizó bajo un sistema convencional de siembra con lámina de agua de 5 cm durante 24 horas. Después de las 48 horas de sembrado se utilizó riego corrido (mojes) cada dos días hasta la primera fertilización. Posteriormente, se dejó una lámina de riego de 5 cm para el control de arvenses. A los 25 días de germinado se dejó constante una lámina de 10 cm y a partir de los 105 días no se llevó a cabo riego alguno.

Tabla 1. Fertilización edáfica aplicada por tratamiento, lote Borriqueros, El Zulia, Norte de Santander

Tratamiento	Numero de abonadas	Dosis total kg/ha abonada	Producto kg/ha									
			Abotain	DAP	KCl	Rafos	Sulfato Fe	Cubo Zinc	Sulfato K	Sulpomag	Nica + B	Llanero ZeO
T1	1	110	31,5	78,5	-	-	-	-	-	-	-	-
	2	192,5	157	-	31,5	-	-	-	-	-	-	-
	3	173	63	-	110	-	-	-	-	-	-	-
	4	110	31	-	78	-	-	-	-	-	-	-
T2	1	181,5	-	-	-	150	10	10	-	-	-	-
	2	175	100	-	-	-	-	-	25	50	-	-
	3	225	100	-	75	-	-	-	-	-	50	-
	4	125	50	-	50	-	-	-	-	25	-	-
T3	1	170	-	-	-	150	10	10	-	-	-	25
	2	175	100	-	-	-	-	-	25	50	-	25
	3	225	100	-	75	-	-	-	-	-	50	-
	4	125	50	-	50	-	-	-	-	25	-	-
T4	1	170	-	-	-	150	10	10	-	-	-	-
	2	175	100	-	-	-	-	-	25	50	-	-
	3	225	100	-	75	-	-	-	-	-	50	-
	4	125	50	-	50	-	-	-	-	25	-	-
T5	1	119,3	31,5	78,5	-	-	-	-	-	-	-	-
	2	188,5	157	-	31,5	-	-	-	-	-	-	-
	3	199,4	63	-	110	-	-	-	-	-	-	-
	4	88	31	-	78	-	-	-	-	-	-	-

Abotain (N 46 %); DAP (N 18 %, P 46 %); KCl (Cl 46 %); Rafos (N 12 %, P 24 %, K 12 %, Mg 2 %, S 1 %, B 0,04 %); Sulfato Fe (Fe 20 %, S 11 %); Cubozinc (N 5 %, Ca 2 %, Mg 6 %, S 9 %, B 2,5 %, Cu 0,5 %, Zn 16 %); Sulfato K (K 50 %, S 18 %); Sulpomag (K 22 %, Mg 18 %, S 21,5 %, Cl 2,5 %); Nica + B (N 15 %, P 4 %, K 4 %); Llanero Zeo (Si 78 %).

Fuente: elaboración propia.

La investigación utilizó un diseño de bloques completos al azar para cinco tratamientos con cuatro repeticiones (cada uno) en un área de 36 m², para un total de 180 m² por tratamiento. Los factores a evaluar en los tratamientos son elementos menores aplicados al suelo (EMS), elementos menores aplicados al follaje (EMF) en combinación con silicio aplicado al suelo (SS), silicio aplicado al follaje (SF) y un adicional que fungirá como testigo (TT), donde se realizó fertilización convencional. Este estudio se realizó considerando cinco tratamientos:

T1: fertilización tradicional con NPK.

T2: fertilización tradicional NPK + aplicación edáfica de elementos menores (Ca, Mg, S, Fe, Mn, Cu, Zn, Cl, B, Mo).

T3: fertilización tradicional NPK + aplicación edáfica de elementos menores (Ca, Mg, S, Fe, Mn, Cu, Zn, Cl, B, Mo) + aplicación edáfica de silicio.

T4: fertilización tradicional NPK + aplicación edáfica de elementos menores (Ca, Mg, S, Fe, Mn, Cu, Zn, Cl, B, Mo) + aplicación foliar de silicio.

T5: fertilización tradicional NPK + aplicación foliar de elementos menores (Ca, Mg, S, Fe, Mn, Cu, Zn, Cl, B, Mo) + aplicación foliar de silicio.

Los datos se corrieron en el sistema de análisis estadístico SAS, versión 9.3, donde se realizó análisis de varianza (ANOVA) y pruebas de Tukey. Los tratamientos 1, 2 y 3 no aparecen en la tabla de dosificación puesto que a estos no se les aplicó fertilización foliar.

Tabla 2. Fertilización foliar aplicada por tratamientos

Tratamiento	N.º de fertilizaciones	Dosis total (l) por ha fertilizada	Producto l/tratamiento	
			Nitrosil-K*	Novaplant**
T4	1	-	-	-
	2	-	-	-
	3	0,5	0,5	-
	4	0,5	0,5	-
T5	1	-	-	-
	2	-	-	-
	3	1	0,5	0,5
	4	1	0,5	0,5

*Nitrosil-K: suministra dosis de silicio al tratamiento.

**Novaplant: suministra elementos menores al tratamiento.

Fuente: elaboración propia.

Calidad molinera del arroz

Para evaluar el componente de molinería (prueba de trilla) se utilizaron 500 gramos de granos por bloque. Las variables de producción fueron: densidad de plantas (plantas/m²), macollamiento (macollas/m²) y número de panículas por área (m²). Dichas variables se determinaron arrojando al azar dentro de cada tratamiento un marco de aforo de un metro cuadrado (1 m²), el cual es elaborado con cuatro varas de 1 metro de largo unidas por sus extremos, formando exactamente un cuadro. Se tomaron 6 panículas al azar por réplica, 24 panículas por tratamientos. Se seleccionó el número de granos totales (NGT) de las panículas para determinar el número de espiguillas. Posteriormente, por panícula se contó el total de granos y estos se dividieron en número de granos llenos (NGLI) y número de granos vanos (NGV) para así determinar el porcentaje de vaneamiento (%Vano), el cual fue calculado a partir de la siguiente fórmula:

$$\text{Vaneamiento} = \left(\frac{\text{Número de granos totales}}{\text{Número de granos vanos}} \right) \times 100$$

Rendimiento del arroz

La muestra representativa de cada bloque estuvo compuesta por seis plantas de cada tratamiento para los componentes de rendimiento, teniendo en cuenta el efecto de borde. Para determinar el rendimiento (kg/ha) se cosechó y se pesó cada tratamiento por separado. La evaluación de los parámetros de calidad molinera involucró el uso de los equipos de laboratorio ubicados en las instalaciones del Molino Oro, propiedad de COAGRONORTE. El parámetro a tener en cuenta dentro de la calidad molinera es el rendimiento de pilada (rendimiento/pilada). Para hallar esta variable es necesario realizar el pesaje de 100 gr de arroz paddy con una humedad entre 11,5 y 12,5 %. Posteriormente, esta muestra es pasada por la descascaradora, para luego ser pesada, pulida y pesada una vez más. Por último, la muestra es analizada en scanner tipo SATAKE RSQI, del cual se obtiene el porcentaje de masa blanca (%MB), grano entero (%GE), grano cristal (%CRIS), harina (%HAR), cascarilla (%CAS) y grano yesado (%GY) del 100 % de la muestra de arroz paddy, esta vez seco.

Tabla 3. Valor F del análisis de varianza para cada variable evaluada

Variable	Plantas/m ²	Macollas/m ²	Pani/m ²	NEspí	NGV	NGLL	NGT	%Vano	kg/ha
GL	4	4	4	4	4	4	4	4	4
F Value	46,27*	59,48	8,53	2,76	19,88	7,88	15,44	15,24	37,19*
Variable	%MB	%GE	%CRIS	%HAR	%CAS	%GY			
GL	4	4	4	4	4	4			
F Value	70,57*	446,9*	176,93*	79*	7,89	82,08*			

* Probabilidad < 0.0001; GL: grados de libertad.

Fuente: elaboración propia.

RESULTADOS Y DISCUSIÓN

El análisis de varianza reportó diferencias altamente significativas entre los tratamientos evaluados ($p < 0,0001$) para las siguientes variables: plantas/m², kg/ha, %MB, %GE, %CRIS, %HAR y %GY (tabla 3).

En la investigación se evaluaron cinco fertilizaciones diferentes enfocadas a la utilización de elementos menores con respecto al plan nutricional normalmente utilizado por los agricultores del sector, llevando a cabo un total de cuatro réplicas del tratamiento en la zona. Se observó que la variable densidad de plantas/m² presentó diferencias significativas ($p < 0,0001$) entre los tratamientos (figura 1A), siendo T3 el que presentó el mayor valor (396,21 plantas/m²), seguido de T5 y T4 (388,16 y 386,81 plantas/m², respectivamente). En contraste T1 (336,34 plantas/m²) y T2 (313,49 plantas/m²) alcanzaron los valores más bajos. Según lo reportado por Ashrafi *et al.* (2014), estos valores manifiestan resultados no significativos debido a que el distanciamiento entre plantas puede verse afectado por la baja cantidad de micronutrientes y el exceso de agua aplicada en los lotes. Por su parte, Ramírez *et al.* (2017) encontraron que al reducir el espaciamiento disminuye el peso de las panículas.

La variable macollamiento/m² presentó diferencias significativas ($p 0,0002$) entre los tratamientos (figura 1B), siendo T4 el de mayor valor (1.221,18 macollas/m²), mientras que T1 reportó el valor más bajo (821,26 macollas/m²). Estos resultados concuerdan con lo reportado por Radhika *et al.* (2013) y Siddika *et al.* (2016), quienes obtuvieron mayor número de macollas productivas en cultivos de arroz tras la aplicación de micronutrientes y silicio dentro de los planes de fertilización.

Así mismo, la variable panículas/m² mostró diferencias significativas ($p 0,0119$) entre tratamientos (figura 1C), siendo T2 (709,09 panículas/m²) y T3 (705,29 panículas/m²) aquellos con los valores más altos. Por otro lado, T5 reportó el menor valor (638,66 panículas/m²). Estos hallazgos fueron mejores a los expuestos por Álvarez *et al.* (2008), Hansch y Mendel (2009), Kumar (2014) y Kumar *et al.* (2015), quienes observaron que la aplicación adecuada de nutrientes es indispensable para obtener cosechas de mayores rendimientos.

La variable número de espiguillas/panícula no presentó diferencias significativas ($p 0,0773$) entre los tratamientos (figura 1D), siendo T5 (10 espiguillas/panícula) el de menor valor y T4 (12,24 espiguillas/panícula) el de mayor valor. Estos resultados contrastan con lo reportado por estudios como los de Yruela (2005), Yamasaki *et al.* (2008), Ahmad *et al.* (2009), Roholla *et al.* (2011), Hafeez *et al.* (2013) y Rout y Sahoo (2015), los cuales manifiestan que la aplicación de microelementos no tiene incidencia en el número de espiguillas por panícula.

De otro lado, la variable número de granos llenos (NGLL) presentó diferencias significativas ($p 0,003$) entre tratamientos (figura 1E), en donde T2 presentó el valor más alto (126,32 granos llenos/panícula) y T1 el más bajo (99,53 granos llenos/panícula). Estos hallazgos coinciden con lo reportado por Álvarez *et al.* (2008), Jin *et al.* (2008), Phattarakul *et al.* (2012) y Yuan *et al.* (2012), quienes expresan que siempre que se aplica silicio aumenta en calidad y peso el número de granos llenos.

La variable número de granos vanos (NGV) también presentó diferencias significativas ($p 0,0003$) entre tratamientos (figura 1F), siendo T1 aquel con el mayor valor (22,28 granos vanos/panícula) y T2 el de menor

valor (10,43 granos vanos/panícula). Estos resultados son similares a los de Álvarez *et al.* (2008), Jin *et al.* (2008), Phattarakul *et al.* (2012) y Yuan *et al.* (2012), puesto que, como señalan estos autores, al aplicar elementos menores y silicio vía edáfica aumenta el número de granos llenos.

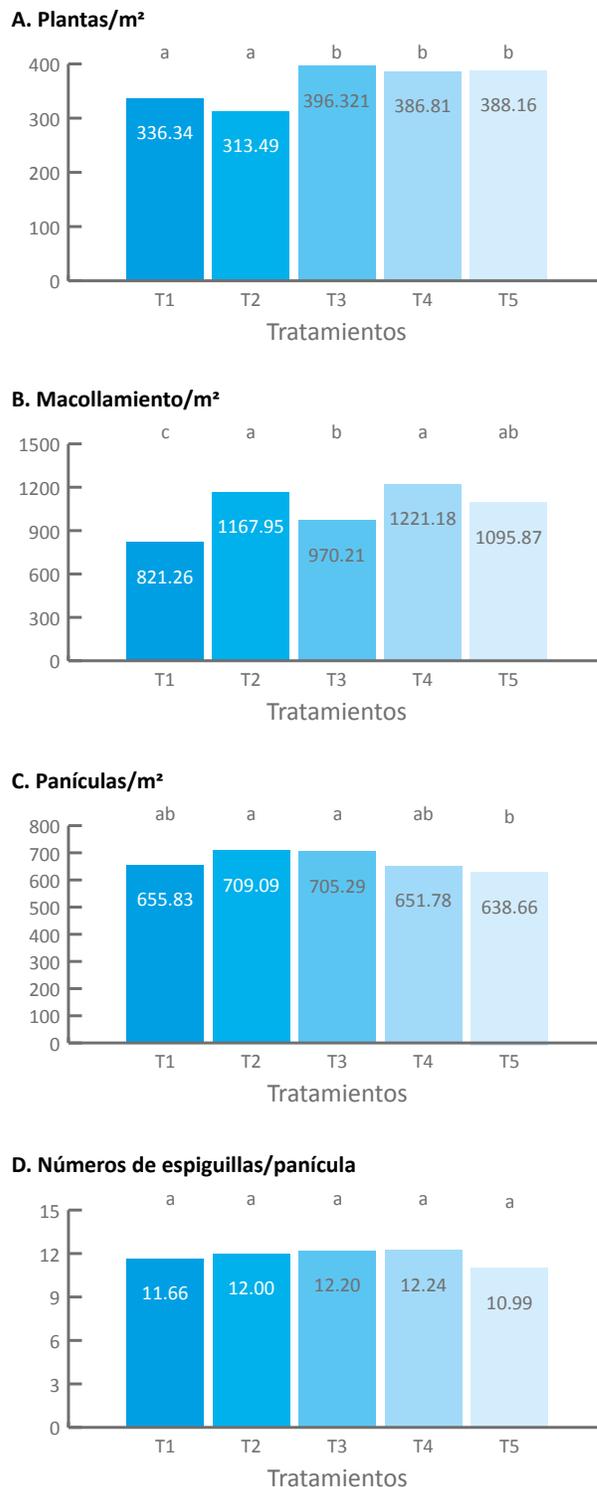
En la misma línea, la variable número de granos totales por panícula (NGT) presentó diferencias significativas ($p < 0.0002$) entre tratamientos (figura 1G), donde T4 (138,77) presentó el mayor número de granos totales por panícula, mientras que T1 (121,84) y T5 (117,53) reportaron los valores más bajos. Los resultados para esta variable concuerdan con lo expresado por Meena y Fathima (2017) y Saha *et al.* (2017), quienes manifiestan que la aplicación de elementos al suelo es importante para la obtención de un mayor número de granos por panícula, aunque esta debe ser complementada con aplicaciones vía foliar.

En cuanto a la variable porcentaje de vaneamiento (%Vano), también se observaron diferencias significativas ($p < 0,0003$) entre los tratamientos (figura 1H). Así, T1 presentó el mayor porcentaje de vaneamiento por panícula (18,48 %) y T2 el menor (7,74 %). Estos resultados se asemejan a lo presentado por Meena y Fathima (2017), Saha *et al.* (2017) y Balindong *et al.* (2018), quienes expresan que la aplicación de fertilizantes al suelo no es suficiente para garantizar el llenado de granos y frutos.

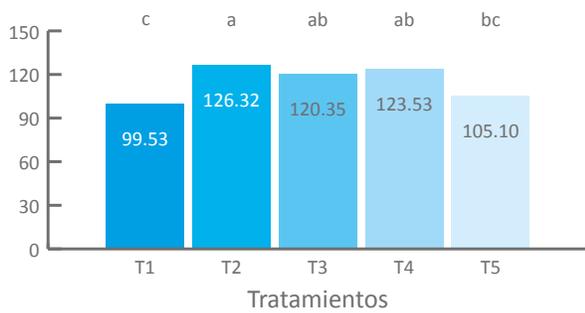
Adicionalmente, se observó que la variable rendimiento (kg/ha) presentó diferencias significativas ($p < 0,0001$) entre tratamientos (figura 1I), siendo T2 el mejor tratamiento en producción (7.689 kg/ha) y T1 el de menor producción (6.886 kg/ha). En el estudio realizado por Álvarez *et al.* (2008), la variable de rendimiento de arroz paddy en la localidad de Ibagué presentó rendimientos de 8.558 kg/ha, mientras que en la localidad de El Guamo esta cifra alcanzó 6.107 kg/ha. Por ende, se puede concluir que el comportamiento de esta variable en la zona de Borriqueros fue aceptable, en comparación con las otras localidades mencionadas. Los resultados obtenidos en esta investigación son similares a los de Shaygany *et al.* (2012), quienes obtuvieron un rendimiento de grano de hasta 6,17 t/ha para fertilización con micronutrientes, en contraste con 4,51 t/ha en el tratamiento control. Del mismo modo, Alamdari y Mobasser (2014) reportaron rendimientos promedios de hasta 4,3 t/ha con la

aplicación de microelementos, en comparación con 2,8 t/ha en el tratamiento testigo.

Figura 1. Análisis de las medias de cada variable



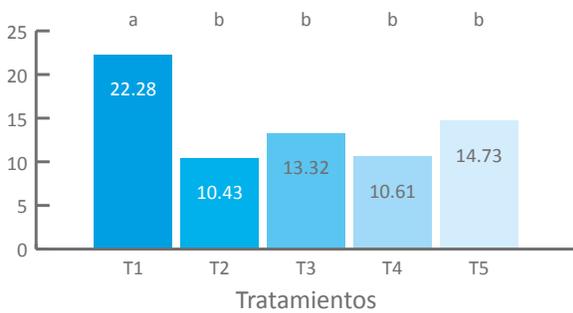
E. Números de granos llenos/panícula



I. Kg/ha



F. Número de granos vanos/panícula



G. Número de granos totales/panícula



H. Vaneamiento



A. Densidad de plantas/m²; B. Macollamiento/m²; C. Panículas/m²; D. Número de espiguillas/panícula; E. Número de granos llenos/panícula; F. Número de granos vanos/panícula; G. Número de granos totales/panícula; H. Porcentaje de vaneamiento; I. Porcentaje de rendimiento (kg/ha).

Fuente: elaboración propia.

Para las variables de producción se tuvo en cuenta los requerimientos arrojados por el análisis PINSAT, obteniendo que el tratamiento con mejor desempeño fue T2 (aplicación de NPK + elementos menores edáficos), cuya dosificación mostró datos importantes para la variable de panículas/m², número de granos llenos/panícula y rendimiento kg/ha. De otro lado, la utilización de un plan nutricional adecuado, como T4 (aplicación de NPK + elementos menores edáficos + silicio foliar), evidenció que la variable de macollamiento/m² y número de granos totales incrementa la producción de arroz. Estos datos coinciden con los obtenidos por Rodríguez-Araujo *et al.* (2010), Shayganya *et al.* (2012), quienes aconsejan hacer aplicaciones de elementos mayores y menores complementadas con aplicaciones foliares, además del uso de silicio como reconstituyente de suelos.

Variables de molinería

La variable porcentaje de masa blanca reportó diferencias significativas ($p < 0,0001$) entre tratamientos (figura 2A), siendo T5 el que obtuvo el mayor porcentaje (70,68 %) y T1 el menor (66,25 %). Estos datos concuerdan con lo presentado por Mejía y Menjivar (2010), Domínguez y Muñoz (2012) y Livore (2013), quienes manifiestan que la realización de evaluaciones de calidad molinera es necesaria para determinar la calidad de las especies de arroz sembradas alrededor del mundo.

Igualmente, la variable porcentaje de granos enteros registró diferencias significativas ($p < 0,0001$) entre tratamientos (figura 2B), siendo T5 el de mayor porcentaje (59,65 %) y T1 aquel con el menor resultado (40,96 %). Estos hallazgos son similares a los reportados por Miranda y Ruíz (2015), quienes expresan que los planes de fertilización que involucran la aplicación de elementos mayores y menores, complementados con la adición de productos foliares, garantizan granos enteros que son bien comercializados.

La variable porcentaje de cristal (grano partido) también registró diferencias significativas ($p < 0,0001$) entre tratamientos (figura 2C), donde T1 obtuvo el mayor porcentaje de granos partidos (25,24 %) y T3 el menor porcentaje (11,09 %). Resultados similares a los de Martínez *et al.* (2010) y Jo y Todorov (2019), quienes manifiestan que el uso de fertilizantes compuestos a base de NPK, complementados con micronutrientes, incrementa de manera notable los componentes del rendimiento y el rendimiento general del cultivo de arroz.

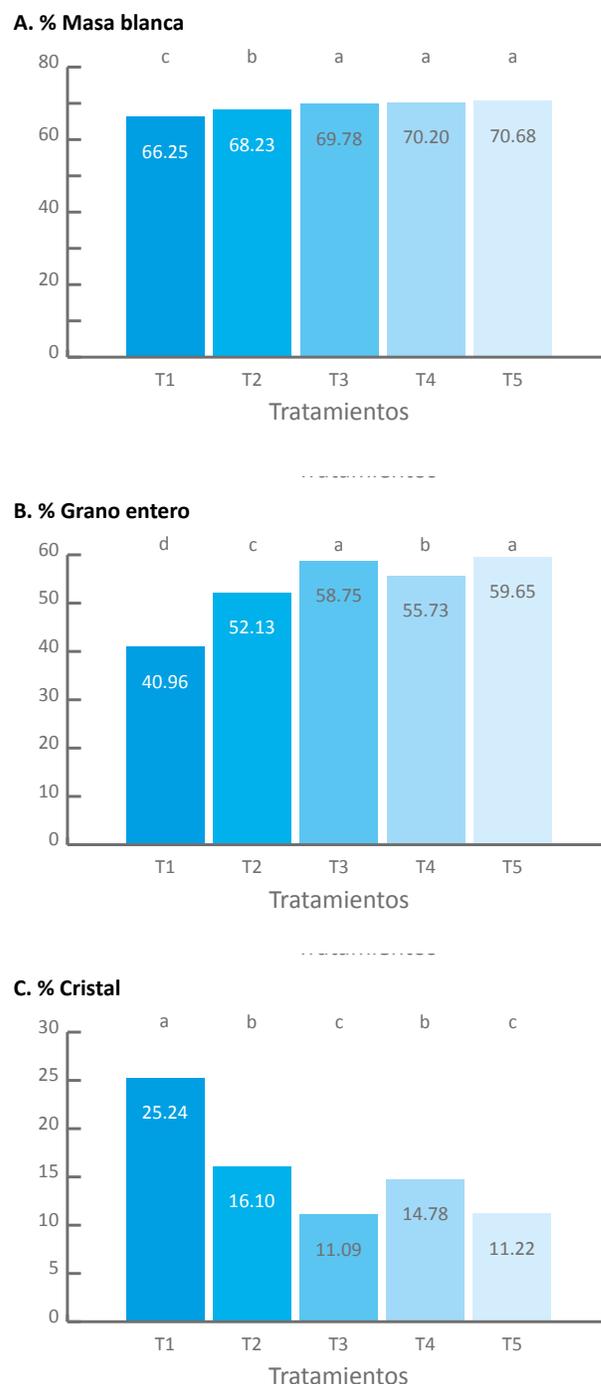
Igual que con las variables anteriores, el porcentaje de harina reportó diferencias significativas ($p < 0,0001$) entre tratamientos (figura 2D), siendo T1 el de mayor porcentaje (12,82 %) y T5 el de la cifra más baja (9,32 %). Estos hallazgos coinciden con los resultados de Larios *et al.* (2010), quienes expresan que la fertilización con micronutrientes no solo mejora notablemente el rendimiento del cultivo, sino que además puede incrementar la calidad nutricional del grano, lo cual resulta importante y atractivo para los consumidores desde la perspectiva de la salud humana, puesto que adquirir un producto biofortificado es clave para contrarrestar problemas de desnutrición.

En la variable porcentaje de cáscara se observaron diferencias significativas ($p 0,0023$) entre tratamientos (figura 2E), donde T1 registró el mayor valor (20,85 %) y T5 el menor porcentaje (19,95 %). Estos resultados son similares a los observados por Siebenmorgen (2013), autor que establece que la fertilización convencional a base de NPK es imprescindible para mejorar la calidad molinera del grano de arroz.

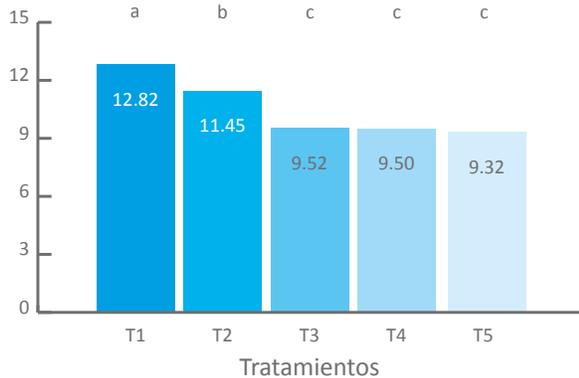
Para la variable porcentaje de yesado también se observaron diferencias significativas ($p < 0,0001$) entre tratamientos (figura 2F), siendo T2 aquel con el mayor porcentaje (22,18 %). Por su parte, el porcentaje de esta variable en los demás tratamientos fue mínimo,

sin superar 1 % de yesado. Esto es similar a lo presentado por Guillermo *et al.* (2012) y Zhang *et al.* (2017), quienes manifestaron que la inclusión de micronutrientes en los planes de fertilización convencional basados en NPK puede incrementar notablemente el rendimiento y la calidad nutraceútica del grano de arroz.

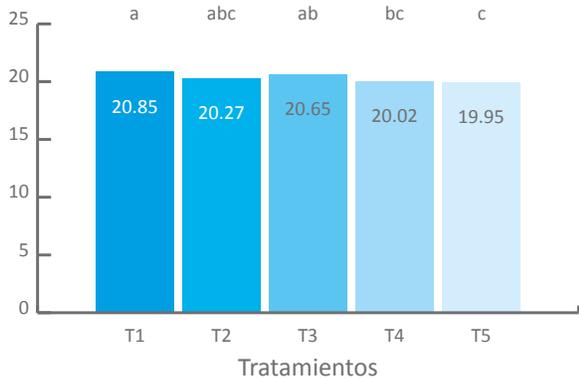
Figura 2. Análisis de las medias de cada variable de molinería



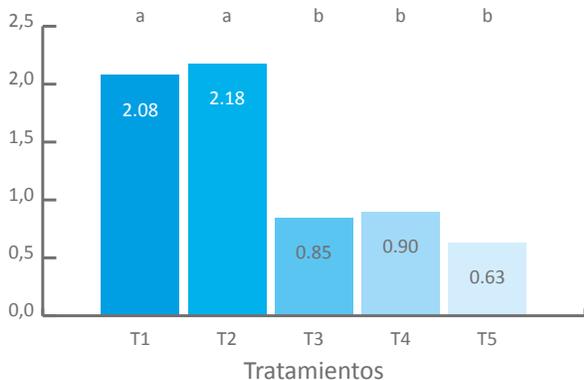
D. % Harina



E. % Cáscara



F. Grano yesado



A. Porcentaje de masa blanca; B. Porcentaje de grano entero; C. Porcentaje de cristal; D. Porcentaje de harina; E. Porcentaje de cáscara; F. Porcentaje de grano yesado.

Fuente: elaboración propia.

Para las variables de calidad molinera el tratamiento con los mejores resultados fue T5 (aplicación de NPK + elementos menores foliares + silicio foliar), el cual demostró buenos efectos sobre las variables porcentaje de masa blanca y porcentaje de grano entero, así como la menor cantidad de granos yesados, cristal y harina. Estos datos que concuerdan con lo expuesto por Mejía y Menjivar (2010), Abubakar *et al.* (2018), quienes expresan que la utilización de fertilización convencional basada en NPK y la adición de elementos menores y silicio mejora la calidad molinera de la variedad Fedearroz 2000.

CONCLUSIONES

Se observó que el tratamiento con mejor comportamiento en cuanto a las variables de producción y rendimiento fue T2 (fertilización NPK + fertilización de elementos menores edáficos), debido a que este mostró un alto grado de aceptación para las variables de panículas/m², número de granos llenos/panícula y rendimiento (kg/ha), en comparación con el testigo (T1), que corresponde al esquema de fertilización más empleado en la zona de Borriqueros.

Se evidenció que a nivel de calidad molinera el tratamiento con mejor comportamiento fue T5 (fertilización NPK + fertilización con elementos menores foliar + fertilización foliar con silicio), el cual permite mejorar los porcentajes de masa blanca y granos enteros, así como obtener el menor porcentaje de granos yesados. Además, la utilización de elementos menores (en especial el silicio) permite fortalecer el sistema inmune de la planta, contribuyendo a combatir agentes bióticos causantes del vaneamiento y comprobando así la importancia de aplicar elementos menores al suelo, los cuales producen una respuesta positiva sobre la calidad molinera del arroz.

La aplicación de fertilizantes NPK + fertilización con elementos menores foliar + fertilización foliar con silicio mejora las condiciones de producción y rendimiento del arroz, lo cual se ve traducido en mejores ingresos económicos para el productor y la entrega de un producto de mejor calidad para el consumo humano.

REFERENCIAS

- Abubakar, B., Yakasai, H. M., Zawawi, N., & Ismail, M. (2018). Compositional analyses of white, brown and germinated forms of popular Malaysian rice to offer insight into the growing diet-related diseases. *Journal of Food and Drug Analysis*, 26(2), 706-715. <https://doi.org/10.1016/j.jfda.2017.06.010>
- Ahmad, W., Niaz, N., Kanwal, S., & Khalid, M. (2009). Role of boron in plant growth: A review. *J Agric Res.*, 47(3), 329-338.
- Alamdari, M., & Mobasser, H. (2014). The effect of macro and micro-nutrient fertilizers on yield and yield attributes of rice in a calcareous soil. *American Journal of Experimental Agriculture*, 4(12), 1604-1615. <https://doi.org/10.9734/ajea/2014/3536>
- Alshaal, T., & El-Ramady, H. (2017). Foliar Application: from Plant Nutrition to Biofortification. *The Environment, Biodiversity & Soil Security*, 1, 71-83. <https://doi.org/10.21608/jenvbs.2017.1089.1006>
- Álvarez, J., Daza, M., & Mendoza, C. (2008). Aplicación de un fertilizante enriquecido con silicio y materia orgánica en arroz (*Oryza sativa* L.) cultivado en Ibagué y el Guamo (Tolima, Colombia). *Revista Facultad Nacional Agronomía Medellín*, 61(2), 4605-4617.
- Ashrafi, E., Pirdashti, H., & Niknejhad, Y. (2014). Effect of iron, zinc and silicon application on quantitative parameters of rice (*Oryza Sativa* L. CV. Tarom Mahalli). *Intl J Farm & Alli Sci.*, 3(5), 529-533.
- Balindong, J. L., Ward, R. M., Liu, L., Rose, T. J., Pallas, L. A., Ovenden, B. W., Snell, P. J., & Waters, D. L. (2018). Rice grain protein composition influences instrumental measures of rice cooking and eating quality. *Rev. Cereal*, 79(18), 35-42. <https://doi.org/10.1016/j.jcs.2017.09.008>
- Barona, E. (2010). Importancia de la semilla de arroz. *Arroz*, 58(489), 15-21.
- Bashir, K., Takahashi, R., Nakanishi, H., & Nishizawa, N. (2013). The road to micronutrient biofortification of rice: Progress and prospects. *Frontiers in Plant Science*, 4(15), 1-7. <https://doi.org/10.3389/fpls.2013.00015>
- Botta, G. F., Tolón-Becerra, A., Lastra-Bravo, X., Hidalgo, R., Rivero, D., & Agnes, D. (2015). Alternatives for handling rice (*Oryza sativa* L.) straw to favor its decomposition in direct sowing systems and their incidence on soil compaction. *Geoderma*, 239-240, 213-222. <https://doi.org/10.1016/j.geoderma.2014.10.021>
- Bouis, H., & Saltzman, A. (2017). Improving nutrition through biofortification: A review of evidence from HarvestPlus, 2003 through 2016. *Global Food Security*, 12, 49-58. <https://doi.org/10.1016/j.gfs.2017.01.009>
- Cakmak, I., & Kutman, U. (2017). Agronomic biofortification of cereals with zinc: A review. *European Journal of Soil Science*, 69(1), 172-180. <https://doi.org/10.1111/ejss.12437>
- Das, J., Salam, R., Kumar, R., & Bhutta, Z. (2013). Micronutrient fortification of food and its impact on woman and child health: A systematic review. *Systematic Reviews*, 2, 67. <https://doi.org/10.1186/2046-4053-2-67>
- Domínguez, A., & Muñoz, O. (2012). Evaluación de la calidad molinera y dimensiones de los granos de dos variedades de arroz y sus varietales de arroz maleza. *Rev. Agronomía Tropical Argentina*, 2(12), 5-6.
- Dutta, S., Das, S., Pale, G., Ingrai, B., Aochen, C., Rai, M., & Pattanayak, A. (2016). Current status and future prospects of research on genetically modified rice: A review. *Agricultural Reviews*, 37(1), 10-18. <https://doi.org/10.18805/ar.v37i1.9259>
- Fairhurst, T. H., & Witt, C. (2002). *Guía práctica para el manejo de nutrientes*. Postash & Phosphate Institute (PPI), Potash & Phosphate Institute of Canada (PPIC), & International Rice Research Institute (IRRI).

- Food and Agriculture Organization of the United Nations [FAO]. (2009). *Global agriculture towards 2050: High Level Expert Forum - How to Feed the World in 2050*. FAO http://www.fao.org/fileadmin/templates/wsfs/docs/Issues_papers/HLEF2050_Global_Agriculture.pdf
- Food and Agriculture Organization of the United Nations [FAO]. (2014). *A regional rice strategy for sustainable food security in Asia and the Pacific*. FAO.
- Gnanamanickam, S. (2009). Rice and its importance to human life. *Prog Biol Con.*, 8, 1-11. https://doi.org/10.1007/978-90-481-2465-7_1
- Guillermo, S., Acosta, C., Cleva, M., & Gonzáles, A. (2012). Sistema de control de calidad de granos de arroz pulidos empleando visión por computadora. *XIV Workshop de Investigadores en Ciencias de la Computación*. Universidad Nacional de Misiones. <http://sedici.unlp.edu.ar/bitstream/handle/10915/18840/Sistema+de+control+de+calidad+de+granos+de+arroz+pulidos.pdf;jsessionid=69E2259FFC4A7D-6F3899E8F1ED4D7C7E?sequence=1>
- Hafeez, B., Khanif, Y., & Saleem, M. (2013). Role of zinc in plant nutrition. A review. *American Journal of Experimental Agriculture*, 3(2), 374-391. <https://doi.org/10.9734/AJEA/2013/2746>
- Hansch, R., & Mendel, R. (2009). Physiological functions of mineral micronutrients (Cu, Zn, Mn, Fe, Ni, Mo, B, Cl). *Current Opinion in Plant Biology*, 12, 259-266. <https://doi.org/10.1016/j.pbi.2009.05.006>
- Hernández-Rodríguez, A., Rives-Rodríguez, N., Acebo-Guerrero, Y., Diaz-de-la-Osa, A., Heydrich-Pérez, M., & Divan-Baldani, L. (2014). Potencialidades de las bacterias diazotróficas asociativas en la promoción del crecimiento vegetal y el control de *Pyricularia oryzae* (Sacc.) en el cultivo del arroz (*Oryza sativa* L.) *Revista de Protección Vegetal*, 29(1), 1-10.
- International Food Policy Research Institute [IFPRI]. (2014). *Global nutrition report 2014: Actions and accountability to accelerate the world's progress on nutrition*. IFPRI. <https://doi.org/10.2499/9780896295803>
- Jin, Z., Minyan, W., Lianghuan, W., Jianguo, W., & Chunhai, S. (2008). Impacts of combination of foliar iron and boron application on iron biofortification and nutritional quality of rice grain. *Journal of Plant Nutrition*, 31, 1599-1611. <https://doi.org/10.1080/01904160802244803>
- Jo, G., & Todorov, T. I. (2019). Distribution of nutrient and toxic elements in brown and polished rice. *Rev Food Chemistry*, 289(2), 360-368. <https://doi.org/10.1016/j.foodchem.2019.03.040>
- Kato, Y., & Katsura, K. (2014). Rice adaptation to aerobic soils: Physiological considerations and implications for agronomy. *Plant Production Sc.*, 17(1), 1-12. <https://doi.org/10.1626/pp.17.1>
- Kumar, S. (2014). Role of micronutrient in rice cultivation and management strategy in organic agriculture. A reappraisal. *Agricultural Sciences*, 5, 765-769. <https://doi.org/10.4236/as.2014.59080>
- Kumar, D., Singh, Sh., Singh, Sw., Mishra, S., Chauhan, D., & Dubey, N. (2015). Micronutrients and their diverse role in agricultural crops: Advances and future prospective. *Acta Physiol Plant.*, 37, 139. <https://doi.org/10.1007/s11738-015-1870-3>
- Larios, A., Porcayo, J., & Poggi, H. (2010). Obtención de una harina de pulido de arroz desengrasada con bajo contenido de fibra neutro detergente. *Rev. Food Chemistry Colombia*, 1(22), 3-5.
- Livore, A. (2013). Calidad industrial y culinaria del arroz. *IDIA XXI*, 6, 190-194
- Martínez, E., Pincioli, M., & Vidal, V. (2010). Proteínas de arroz propiedades estructurales y funcionales. *Rev Argentina*, 2(2), 67-69.
- Meena, N., & Fathima, P. (2017). Nutrient uptake of rice as influenced by agronomic biofortification of Zn and Fe under methods of rice cultivation. *Int J Pure App Biosci.*, 5(5), 456-459. <https://doi.org/10.18782/2320-7051.3027>
- Mejía, M. S., & Menjivar, J. C. (2010). Nutrición mineral de arroz. En V. M. Degiovanni-Beltramo, C. P. Martínez-Racines, & F. Motta (eds.). *Producción eco-eficiente del arroz en América Latina*. Centro Internacional de Agricultura Tropical (CIAT).

- Mercado, S., Caleño, J., & Rozo, L. (2020). Improvement of the methodology of the tetrazolium test using different pretreatments in seeds of the genus *Epidendrum* (Orchidaceae). *Journal of Seed Science*, 42, e202042013. <https://doi.org/10.1590/2317-1545v42231028>
- Miranda, E., & Ruíz, R. (2015). Análisis proximal de granos de arroz, frijol, maíz y café comercializados en el mercado. *Revista Roberto Huembes de Managua*, 6(18), 2-9.
- Mohanty, S. (2013). Trends in global rice consumption. *Rice Today*, 12(1), 44-45.
- Muthayya, S., Sugimoto, J., Montgomery, S., & Maberly, G. (2014). An overview of global rice production, supply, trade, and consumption. *Ann N. Y. Acad Sci*, 1324, 7-14. <https://doi.org/10.1111/nyas.12540>
- Phattarakul, N., Rerkasem, B., Li, L., Hu, L., Zou, C., Ram, H., Sohu, B., Kang, B., Surek, H., Kalayci, M., Yazici, A., Zhang, F., & Cakmak, I. (2012). Biofortification of rice grain with zinc through zinc fertilization in different countries. *Plant Soil*, 361, 131-141. <https://doi.org/10.1007/s11104-012-1211-x>
- Radhika, K., Hemalatha, S., Maragatham, S., & Praveena, S. (2013). Effect of foliar application of micronutrients on the yield components of rice and soil available micronutrients status. *Asian Journal of Soil Science*, 8(2), 419-421.
- Ramírez, E., Caraballo, R., & Medina, V. (2017). *Efecto de la distancia entre plantas y la época de siembra en las características morfológicas de la panícula de la planta de arroz (O. sativa L.)*. Instituto de Investigaciones del Arroz.
- Rives, N., Acebo, Y., & Hernández, A. (2007). Bacterias promotoras del crecimiento vegetal en el cultivo del arroz (*Oryza sativa* L.). Perspectivas de su uso en Cuba. *Cultivos Tropicales*, 28(2), 29-38.
- Rodríguez-Araujo, E. A., Bolaños-Benavides, M. M., & Menjivar-Flores, J. C. (2010). Efecto de la fertilización en la nutrición y rendimiento de ají (*Capsicum* spp.) en el Valle del Cauca, Colombia. *Acta Agronómica*, 59(1), 55-64. <https://www.redalyc.org/articulo.oa?id=1699/169916223005>
- Roholla, S., Shahsavari, M., & Rezaei, M. (2011). A general overview on manganese (Mn) importance for crops production. *Australian Journal of Basic and Applied Sciences*, 5(9), 1799-1803.
- Rout, G., & Sahoo, S. (2015). Role of iron in plant growth and metabolism. *Reviews in Agricultural Science*, 3, 1-24. <https://doi.org/10.7831/ras.3.1>
- Saha, S., Chakraborty, M., Padhan, D., Saha, B., Murmu, S., Batabyal, K., Seth, A., Hazra, G., Mandal, B., & Bell, R. (2017). Agronomic biofortification of zinc in rice: Influence of cultivars and zinc application methods on grain yield and zinc bioavailability. *Field Crops Research*, 210, 52-60. <https://doi.org/10.1016/j.fcr.2017.05.023>
- Salazar, S. A., Botello, H., & Quintero, J. (2020a). Optimización de la prueba de tetrazolio para evaluar la viabilidad en semillas de *Solanum lycopersicum* L. *Ciencia & Tecnología Agropecuaria*, 21(3). https://doi.org/10.21930/rcta.vol21_num3_art:1344
- Salazar, S. A., Quintero, Rojas, J. (2020b). Optimization of the tetrazolium test in three species of orchids of the Andean forest. *Australian Journal Crop Sci*, 14(5), 822-830. <https://doi.org/10.21475/ajcs.20.14.05.p2276>
- Salazar-Mercado, S., Quintero-Caleño, Q., & Rojas-Suárez, J. (2020) Determinación de la viabilidad de semillas de *Capsicum annum* L. utilizando la prueba de tetrazolio. *Aibi Revista Investig Adm Ing.*, 8(3), 7-12. <https://doi.org/10.15649/2346030X.718>
- Serraj, R., McNally, K., Slamet, I., Kohli, A., Haefele, S., Atlin, G., & Kumar, A. (2011). Drought resistance improvement in rice: An integrated genetic and resource management strategy. *Plant Production Science*, 14(1), 1-14. <https://doi.org/10.1626/pp.14.1>

Shayganya, J., Peivandya, N., & Ghasemi, S. (2012). Increased yield of direct seeded rice (*Oryza sativa* L.) by foliar fertilization through multi-component fertilizers. *Archives of Agronomy and Soil Science*, 58(10), 1091-1098. <https://doi.org/10.1080/03650340.2011.570336>

Siddika, M., Abedin, M., Sharmin, T., Hanif, M., & Chandra, P. (2016). Effect of different micronutrients on growth and yield of rice. *International Journal of Plant and Soil Science*, 12(6), 1-8. <https://doi.org/10.9734/IJPSS/2016/28707>

Siebenmorgen, T. (2013). Laboratory measurement of rice milling yield. En J. T. Hardke (ed.), *Rice production handbook* (pp. 167-173). University of Arkansas.

Virgili, A. (1996). Introducción a la fertilización con microelementos. *Welgro Micromix*, 1(2), 3-10.

Yamasaki, H., Pilon, M., & Shikanai, T. (2008). How do plants respond to copper deficiency?. *Plant Signaling & Behavior*, 3(4), 231-232. <https://doi.org/10.4161/psb.3.4.5094>

Yuan, L., Wu, L., Yang, C., & Lv, Q. (2012). Effects of iron and zinc foliar applications on rice plants and their grain accumulation and grain nutritional quality. *J Sci Food Agric.*, 93(2), 254-261. <https://doi.org/10.1002/jsfa.5749>

Yruela, I. (2005). Copper in plants. *Brazilian Journal of Plant Physiology*, 17(1), 145-156. <https://doi.org/10.1590/S1677-04202005000100012>

Zhang, C., Zhao, W., Gao, A., Su, T., Wang, Y., Zhang, Y., Zhou, X., & He, X. (2017). How could agronomic biofortification of rice be an alternative strategy with higher cost-effectiveness for human iron and zinc deficiency in China?. *Food and Nutrition Bulletin*, 39(2), 246-259. <https://doi.org/10.1177/0379572117745661>

Análisis multicriterio de plantas medicinales cultivadas en la vereda Ceylan del municipio de Viotá, Cundinamarca, Colombia

Multi-criteria Analysis of the Medicinal Plants Grown in Ceylan, municipality of Viotá, Cundinamarca, Colombia

Yuri Alicia Chávez-Plazas^{ac}, María Lucero Ramírez-Mahecha^{ad}, Judith Elena Camacho-Kurmen^{be}

^a Grupo de investigación ODISEA, Facultad Ciencias Sociales, Universidad Colegio Mayor de Cundinamarca, Colombia

^b Grupo de investigación Bioprocesos y Control (GBYC), Facultad Ciencias de la Salud, Universidad Colegio Mayor de Cundinamarca, Colombia

^c yachavez@unicolmayor.edu.co | <https://orcid.org/0000-0003-0115-7010>

^d mluceroramirez@unicolmayor.edu.co | <https://orcid.org/0000-0002-3378-6033>

^e jelenacamacho@unicolmayor.edu.co | <https://orcid.org/0000-0002-8880-1501>

RESUMEN

El presente trabajo se orientó a realizar el análisis multicriterio de plantas medicinales cultivadas en la vereda Altos de Ceylan, municipio de Viotá (Cundinamarca, Colombia), como un paso hacia la certificación ecológica y el aprovechamiento sostenible de la biodiversidad existente en el país, combinados con el diálogo de saberes, con el fin de rescatar el conocimiento tradicional de la comunidad y construir nuevos conocimientos. El análisis multicriterio define los criterios ambientales y económicos que se ponderaron para la selección de las plantas medicinales a certificar como productos ecológicos. “El diálogo de saberes” es una metodología que establece un intercambio de saberes entre la academia —en este caso la Universidad Colegio Mayor de Cundinamarca y la Universidad

Nacional de Colombia—, los investigadores y los conocimientos de las comunidades locales, representadas en esta investigación por las mujeres rurales pertenecientes a la Asociación Semillas de Esperanza y Paz de Mujeres Víctimas del Conflicto Armado (Asepamuvic). El conocimiento construido versó sobre los saberes tradicionales con respecto al cultivo y el manejo y uso de plantas medicinales y aromáticas cultivadas en la vereda, identificando como especies promisorias a la sábila (*Aloe vera* Linneo), el limoncillo (*Cymbopogon citratus*), el prontoalivio (*Lippia alba*-Verbenaceae), la manzanilla (*Matricaria chamomilla* Linneo), la

Citation: Chávez-Plazas, Y. A., Ramírez-Mahecha, M. C. y Camacho-Kurmen, J. E. (2021). Análisis multicriterio de plantas medicinales cultivadas en la vereda Ceylan del municipio de Viotá, Cundinamarca, Colombia. *Mutis*, 11(1), 22-36.
<https://doi.org/10.21789/22561498.1751>

Recibido: Abril 6, 2021. **Aceptado:** Mayo 13, 2021.

Copyright: ©2021. Chávez-Plazas, Y. A., Ramírez-Mahecha, M. C. y Camacho-Kurmen, J. E. (2021). This is an open-access article, which permits unrestricted use, distributions and reproduction in any medium, provided the original author and source are credited.

Competing Interests: The authors have no conflict of interest.



hierbabuena (*Mentha piperita* Linneo), el tomillo (*Pectis graveolens* Klatt), el perejil (*Petroselinum crispum* (Mill.)) y el romero (*Rosmarinus officinalis* Linneo).

Palabras clave: biodiversidad, economía ambiental, intercambio de conocimientos, medicina tradicional.

ABSTRACT

This work carried out a multi-criteria analysis of the medicinal plants grown in the village “Altos de Ceylan,” municipality of Viotá (Cundinamarca, Colombia), as a step towards the ecological certification and sustainable use of the existing plant biodiversity in Colombia, combined with the dialogue of knowledge, which is aimed at rescuing traditional knowledge by local communities and building new insights. The multi-criteria analysis allowed us to define the environmental and economic criteria weighted for the selection of the medicinal plants to be certified as ecological products. “The dialogue of knowledge” is a methodology that establishes an exchange of knowledge between the academy—in this case, Universidad Colegio Mayor de Cundinamarca and the National University of Colombia—, researchers and local communities, represented in this study by women belonging to “Seeds of Hope and Peace, Association of Women Victims of Armed Conflict” [*Asociación Semillas de Esperanza y Paz de Mujeres Víctimas del Conflicto Armado*] (Asepamuvic). The insights built encompassed traditional knowledge regarding the cultivation, management and use of locally-grown medicinal and aromatic plants, making it possible to identify promising species such as sábila (*Aloe vera* Linnaeus), lemongrass (*Cymbopogon citratus*), prontoalivio (*Lippia alba-Verbena*), chamomile (*Matricaria chamomilla* Linneo), mint (*Mentha piperita* Linneo), thyme (*Pectis graveolens* Klatt), parsley (*Petroselinum crispum* (Mill.)), and rosemary (*Rosmarinus officinalis* Linneo), based on the analysis of environmental and economic criteria.

Keywords: Biodiversity, environmental economics, knowledge exchange, traditional medicine.

INTRODUCCIÓN

La Organización Mundial de la Salud (OMS) ha defendido la utilidad y conservación de los saberes tradicionales desde la medicina tradicional, formulando periódicamente, desde 2002, los lineamientos estratégicos de medicina tradicional para el mundo. Este ente internacional resalta además la importancia de conservar plantas medicinales para su uso desde la práctica de atención prioritaria en salud, como quedó establecido en la Declaración de Chiang Mai de 1988 (OMS, 1988). Por lo anterior, este trabajo se llevó a cabo de manera conjunta con las comunidades rurales de la vereda de Altos Ceylan, particularmente con la Asociación Semillas de Esperanza y Paz de Mujeres Víctimas del Conflicto Armado (Asepamuvic), desde lo estipulado en la Política Nacional Para la Gestión Integral de la Biodiversidad y sus Servicios Ecosistémicos, donde se establece la importancia de conocer, preservar, restaurar, conservar y usar sosteniblemente la biodiversidad para identificar especies promisorias de plantas nativas (MADS, 2012).

El diálogo de saberes hace parte de la metodología de investigación-acción participativa (IAP), la cual se basa en la interacción entre distintos actores para el reconocimiento de los problemas propios de su territorio u organización. Este abordaje permite plantear alternativas de mejoramiento al problema estudiado, para lo cual se requiere de la participación de la comunidad y de la generación de aprendizaje mutuo durante todo el proceso (Delgado & Rist, 2016; Ochoa, 2012).

El diálogo de saberes es importante para encontrar puntos en común que permitan validar y fortalecer la experiencia de los actores y los espacios que estos intervienen, con el fin de consolidar un marco conceptual a partir de los encuentros, las reflexiones y las estrategias de estos actores para atender las necesidades, los acuerdos y los enfoques establecidos, acogiendo así el llamado de la interculturalidad como proceso vigente dentro del desarrollo de los estudios sociales y comunitarios (Delgado & Rist, 2016; Ochoa, 2012; Romero & Rodríguez, 2010). Adicionalmente, desde otra perspectiva, el diálogo intercultural permite que las comunidades se sientan respetadas e involucradas en el tránsito a través del conocimiento continuo de la enfermedad

hacia la salud (Delgado & Rist, 2016; Hernández-Rincón *et al.*, 2017; Ochoa, 2012).

De acuerdo con cifras reportadas por el Departamento Administrativo Nacional de Estadística (DANE), solo durante 2017 las ventas al exterior de hierbas aromáticas lograron un crecimiento de 18 %, alcanzando un total de USD 37,2 millones (La República, 2018). Además, en 2018 Colombia exportó 101.054 toneladas de plantas medicinales y aromáticas, siendo Estados Unidos, Canadá y la Unión Europea los principales destinos. Las variedades que más se venden al extranjero son albahaca, hierbabuena, tomillo y romero (MADR, 2019).

Los productos naturales y las plantas medicinales en Colombia cuentan con estándares de calidad que los garantizan, además de la vigilancia del Ministerio de Salud y el Instituto Nacional de Vigilancia en Medicamentos y Alimentos (Invima), siendo este último el encargado de otorgar los registros sanitarios, establecer las buenas prácticas de manufactura y regular la comercialización y producción de este tipo de productos por medio de la “lista básica”, en la cual se encuentran las plantas medicinales aptas para su uso y consumo, siendo estas en mayor proporción especies no nativas (Castellanos *et al.*, 2017; Invima, 2019).

El uso comercial de plantas medicinales genera beneficios monetarios y no monetarios, dentro de los cuales se cuentan el conocimiento, la conservación y el uso sostenible de especies promisorias, tanto a nivel medicinal como industrial, la creciente tendencia del mercado a exigir calidad en las materias primas y la eficiencia en el proceso productivo para la obtención de la planta medicinal en términos de volumen, continuidad y calidad. Además de lo anterior, se tienen como beneficios el valor agregado que se genera al recuperar conocimientos tradicionales y utilizar la bioprospección; la capacitación que se brinda a cultivadores, recolectores y compradores, resaltando la importancia de los valores locales e individuales, así como el valor del medio ambiente; y la generación y el establecimiento de negocios viables y sustentables, que hacen posible generar desarrollo local, regional y nacional a través del uso comercial de plantas medicinales de manera sustentable (Kate & Laird, 2019).

Con base en lo anterior, esta investigación tiene como objetivo realizar un análisis multicriterio (AMC) en el que se ponderan las categorías y los criterios de selección ambientales y económicos establecidos por

Camacho y Castiblanco (2008) para establecer las plantas medicinales y aromáticas a incorporar dentro del proceso de certificación ecológica de las especies cultivadas en la vereda Altos de Ceylan. Para ello, este trabajo adopta diferentes estrategias formuladas desde el diálogo de saberes, las cuales permiten la integración del conocimiento popular y científico (Reyes-Sánchez, 2009). El AMC propuesto se basa en la ponderación y compensación de variables, ya sean determinantes o factores de aptitud, exigiendo la indicación de pesos o ponderaciones para cada factor (Sipahi & Mehpare, 2010).

MATERIALES Y MÉTODOS

El desarrollo de esta investigación acerca de la aplicación de AMC para el proceso de certificación de plantas medicinales y aromáticas, que se da en el contexto del proyecto “Diálogo de saberes: una experiencia investigativa con mujeres rurales de Viotá orientada a la certificación ecológica en plantas medicinales y aromáticas”, tiene en cuenta las fases del proceso de diálogo de saberes desarrollado con mujeres rurales de Viotá asociadas a Asepamuvic, para lo cual se aplicó la metodología propuesta por Morse (1994), cuyas fases serán descritas a continuación.

Fase de reflexión

Esta fase consiste en la identificación del tema, la pregunta de investigación y la perspectiva metodológica. Durante esta fase se utilizó la metodología del tendero social, la cual es ampliamente usada en entornos colaborativos. Esta metodología proporciona un espacio físico para que los participantes presenten sus resultados (escritos o dibujados), después de la realización de las dinámicas propuestas, con el fin de generar un registro visible de lo que sucede en una comunidad (BBVA, 2017). Se realizaron talleres de capacitación e intercambio de experiencias con 15 mujeres rurales integradas a Asepamuvic y, eventualmente, dos hombres de la vereda, teniendo en cuenta el conocimiento tradicional de los participantes y el conocimiento científico construido en la universidad referente a formas de cultivo. Esto permitió identificar los usos de las plantas, las medidas de bioseguridad requeridas para su manejo y las buenas prácticas recomendadas para su cultivo. Así mismo, fue posible resolver preguntas tales como ¿qué plantas reconozco?,

¿qué usos y beneficios tienen estas?, ¿cuáles son los métodos de cultivo tradicionales y ecológicos?, ¿cuál sería la viabilidad económica de tales métodos?

Fase de planificación

En esta fase se llevó a cabo la selección del contexto, las estrategias y la propuesta de investigación (visitas previas, reconocimiento del contexto y de la población). Como estrategia, se trabajó por grupos, donde cada uno de estos estableció las plantas aromáticas que cultivaría y la técnica de cultivo a emplear a partir de la construcción de un protocolo de cultivo. El objetivo de esta fase fue significar el conocimiento de las comunidades y complementarlo con el conocimiento construido desde la universidad. Esta primera actividad finalizó con una visita a los cultivos tradicionales de los participantes.

Fase de entrada

Se seleccionaron los informantes y los casos. De igual forma, se inició la realización de entrevistas y observaciones (planificación de los posibles escenarios de los diálogos). Además, se elaboraron herbarios con las plantas medicinales y aromáticas cultivadas en la vereda, en los que se incluyeron el nombre de cada planta, sus usos y preparaciones. Lo anterior tuvo como objetivo rescatar el conocimiento tradicional sobre las plantas en la vereda Altos de Ceylan. El trabajo se desarrolló en tres etapas: proyecto productivo, trabajo en equipo y aprovechamiento de la biodiversidad con respecto a plantas medicinales y aromáticas (conocer, conservar, restaurar y utilizar) (MADS, 2012).

La muestra de participantes estuvo conformada por 13 mujeres y 2 hombres residentes de la vereda, en edades comprendidas entre 18 y 75 años. Varios de los participantes se dedican al cultivo de frutales y café (*Senna hirsuta*), así como a la cría y venta de porcinos. La mayoría de estos cuenta con nivel de educación primario, mientras que un pequeño grupo no ha recibido educación formal. Los participantes emplean sus huertas caseras para el cultivo de plantas medicinales.

Fase de recolección reproductiva de la información

En esta etapa se realizó una primera devolución de la información construida mediante el diálogo de

saberes a los participantes. Para ello, se retomó la información de los herbarios, la cual fue complementada con información científica ampliada frente a los usos de cada planta, su nombre científico, usos y propiedades. Esto permitió reforzar el importante recurso que significa cada planta identificada.

Fase de salida de campo

Esta fase se realizó en tres momentos. En primer lugar, se visitaron los cultivos de plantas medicinales en las huertas de los participantes. El segundo momento contempló la realización de una exposición sobre las características del cultivo y la aplicación de buenas prácticas agrícolas (BPA) por parte de cada grupo de participantes, actividad que fue retroalimentada por los investigadores. El tercer momento fue la visita del Instituto Agroecológico María Cano al proyecto, la cual tuvo como objetivo fortalecer la elaboración de productos derivados de las plantas cultivadas y seleccionadas.

Análisis multicriterio

Se utilizó la metodología de análisis multicriterio (AMC) desarrollada por Camacho y Castiblanco (2008) para ponderar y seleccionar las plantas medicinales que serían incluidas en el programa de certificación ecológica, luego de comprobar que dichas plantas fueron cultivadas de manera orgánica, sin el uso de agroquímicos (MADR, 2019).

Diseño experimental y análisis de datos

Se realizó un análisis multicriterio en el que se asignó un valor a los criterios ambientales y económicos considerados, como se muestra a continuación.

Criterio ambiental

Especies nativas

En Colombia, el universo de plantas con actividad medicinal es amplio (156 especies comercializadas). Por ende, se tuvo en cuenta el criterio de que la planta medicinal fuese una especie nativa para así aprovechar la biodiversidad del país. Además, se consideró la frecuencia de comercialización de dicha especie, así como la demanda de los laboratorios naturistas y su inclusión en el listado de plantas medicinales y aromáticas comercializadas en Colombia (Bernal *et al.*, 2011;

Castellanos *et al.*, 2017; Invima, 2019; Rincón-Parra, 2012). Así, se consideran los siguientes criterios:

- Criterio: nativa = 100 %.
- Criterio: no nativa = 0 % (naturalizada [adaptada a nuestros ecosistemas], foránea o importada).

Seguridad del uso de la especie (calidad, efectividad e inocuidad)

Para cada una de las especies seleccionadas se realizó un estudio bibliográfico tomando en cuenta aspectos relacionados con la seguridad del uso de la especie, lo cual hace referencia a que sean plantas efectivas (con acción farmacológica), que cumplan con las normas de calidad de las farmacopeas y que sean inocuas, es decir, sin antecedentes de toxicidad. Al respecto, se tiene los siguientes criterios:

- Estudios farmacológicos, toxicológicos y clínicos = 50 %.
- Inclusión en lista básica del Invima = 50 %.
- No se conocen estudios = 0 %.
- No inclusión en la lista básica del Invima = 0 %.

Aprovechamiento de la biodiversidad

En lo relacionado con el aprovechamiento de la biodiversidad existente, se cuentan los siguientes criterios:

- Participación en estudios de bioprospección = 50 %.
- Conocimiento tradicional documentado = 50 %.
- Sin información = 0 %.

Para seleccionar los criterios antes descritos se toma como referente el objetivo principal de la Política Nacional de Biodiversidad, que es conocer, conservar y hacer uso sostenible de la biodiversidad a través de acciones como recuperar y divulgar el conocimiento y las prácticas tradicionales y hacer uso de la bioprospección para financiar la conservación de la biodiversidad (MADS, 2012).

Fuentes, sustentabilidad y calidad de la materia prima

La sustentabilidad se relaciona con el número de especies promisorias y la forma en la que se obtiene la planta medicinal (recolección silvestre o cultivo). Como criterio se tiene:

- Mayor sustentabilidad = 10 puntos.
- Más o menos sustentable = 5 puntos.
- Menor posibilidad de sustentabilidad = 2 puntos.

La manera en la que se obtiene la planta medicinal tiene implicaciones importantes para la sustentabilidad del proceso, por lo que es importante diferenciar si la obtención de la planta se hace por recolección silvestre o por cultivo, puesto que a partir del método de obtención se ven afectados factores como la utilidad, la fluctuación de oferta, el control de calidad, la identificación botánica, el mejoramiento genético, la manipulación agronómica, el manejo postcosecha y la posible adulteración de una especie determinada (Kate & Laird, 2019).

Criterios económicos

Valor económico total (VET)

Para este análisis se tuvo en cuenta el concepto de valor económico total (VET) planteado por David Pearce (Pearce & Turner, 1995), el cual hace referencia a todos los posibles valores de un bien ambiental, independientemente de que estos tengan precios de mercado o no.

Desde esta perspectiva, se consideran los valores de uso (directo e indirecto) y de no uso (valor de opción y valor de existencia) de las plantas medicinales, puesto que la contribución porcentual de las especies y los ecosistemas silvestres a la economía de los países en desarrollo, como Colombia, suele ser mucho mayor, por lo cual se toma como referente la mencionada valoración (tabla 1).

Tabla 1. Categorías de valor incorporadas al análisis de valor económico total de los bienes de un ecosistema

Valor de uso		Valor de no uso	
Valor de uso directo	Valor de uso indirecto	Valor de opción	Valor de existencia
Alimentos vegetales = 1	Retención de sedimentos = 1	Especies = 1	Especies en extinción = 1
Farmacéuticos = 1	Retención nutrientes = 1	Conservación de hábitat = 1	Estética = 1
Materia prima = 1	Soporte a biodiversidad = 1	Protección de la biodiversidad = 1	Conservación = 1
Investigación = 1	Producción de O ₂ = 1	Potencial farmacéutico = 1	-
Reproducción de especies = 1	Secuestro CO ₂ = 1	-	-
Biomasa = 1	Belleza escénica = 1	-	-
Plantas medicinales = 1	Polinización = 1	-	-
Plantas ornamentales = 1	Reproducción de especies = 1	-	-
Subtotal = 8	Subtotal = 8	Subtotal = 4	Subtotal = 3
Total = 23			

Fuente: Barzev (2002).

Como criterio, se toma cada categoría como variable dicotómica, de la siguiente manera:

- No aporta al valor = 0.
- Aporta al valor = 1.
- Valor total (valor de uso + valor de no uso) = 23.

Muchos ecosistemas naturales reportan beneficios de carácter indirecto y su valor económico deriva de los servicios que proveen, más que de los productos que se pueden obtener de ellos. Por ende, se tuvo en cuenta tanto el valor de uso como el de no uso, ya que, por ejemplo, muchas plantas están dotadas de atributos socioculturales y religiosos que les otorgan una condición especial. Por su parte, el estudio del carácter físico o biológico de las plantas ha redundado en la acumulación de conocimientos científicos básicos de utilidad, que a su vez ha sentado las bases para nuevos adelantos y nuevos descubrimientos (Jiménez-Herrero, 1997).

Otros criterios económicos

Dentro de este conjunto de criterios se tiene en cuenta el valor agregado derivado de la recuperación del conocimiento tradicional y su utilización en la bioprospección. Así mismo, se incluyen la capacitación a cultivadores, recolectores y compradores —proceso en el que se resaltan los valores locales, individuales y del medio ambiente— y la generación y

el establecimiento de negocios viables y sustentables, con lo cual se genera desarrollo local, regional y nacional a partir del uso comercial de plantas medicinales desde una perspectiva sustentable en términos económicos, sociales y ecológicos.

Dentro del grupo de otros criterios económicos se incluyeron las especies promisorias, las necesidades del mercado, la eficiencia, el valor agregado, el desarrollo local y la capacitación, asignado un máximo de 10 puntos por categoría.

Necesidades del mercado

Este criterio hace referencia a la calidad del producto. Al ser plantas ampliamente comercializadas en Colombia, y con potencial a nivel internacional, surge la necesidad de incluir esta variable, puesto que a partir de criterios de calidad y del conocimiento de la especie medicinal correcta será posible cubrir el mercado local en cuanto a volumen (cantidades), eventualmente abriendo paso a la posibilidad de exportar estos productos.

Eficiencia

La sostenibilidad (eficiencia) se logra cuando se cuenta con cosechas productivas, se conservan los recursos (especies, biodiversidad, agua y suelo), se satisfacen las necesidades económicas y sociales, se evidencian sistemas viables de producción, se valora

el conocimiento tradicional y la información y se registran bajos niveles de plagas y enfermedades. Otros determinantes de la eficiencia son la educación, la investigación, la difusión y el desarrollo de la agricultura internacional.

Valor agregado

Para determinar el valor agregado de los productos se tiene en cuenta que estos sean cosechados por comunidades locales, aprovechando el conocimiento y las capacidades de la mujer, quien tiene un papel primordial en el cultivo, las familias y la comunidad, generando así beneficios sociales, empleo y una economía de subsistencia para la región. Además, en esta dimensión se valora el conocimiento tradicional asociado al desarrollo y los métodos tradicionales de cultivo, así como los estudios de investigación y de bioprospección realizados.

Capacitación y desarrollo local

Se debe tener en cuenta la importancia de la capacitación en buenas prácticas agrícolas y de recolección para todos los actores que intervienen en el proceso de producción de plantas medicinales, situación que redundará en desarrollo local y regional en forma sostenible.

RESULTADOS

Los participantes de esta investigación pertenecen a la asociación Asepamuvic. La edad promedio de los participantes fue 36 años. El 98 % de los asistentes son mujeres y el 2 % restante corresponde a hombres. El nivel de escolaridad de los participantes está entre educación primaria y secundaria (67 %), técnica (13 %) y universitaria (7 %), encontrando una tasa de analfabetismo de 13 %. El 4 % de los participantes son adultos mayores con gran conocimiento tradicional sobre el uso de plantas medicinales y aromáticas.

El listado de plantas medicinales que se encuentran presentes y se cultivan en la vereda Altos de Ceylan, Viotá, Cundinamarca, se construyó teniendo en cuenta el nombre tradicional y científico de cada una de ellas, así como sus usos, preparaciones y propiedades. Las plantas incluidas en este listado son: limonaria (*Cymbopogon citratus*), hierba del dolor (*Acalypha arvensis*), poleo (*Minthostachys tomentosa*), lengua de suegra (*Sansevieria trifasciata* Prain), perejil

(*Petroselinum crispum* (Mill.)), borrachero (*Atropa belladonna* Linneo), guaba (*Phytolacca icosandra*), comfrey (*Symphytum officinale* Linneo), sauco (*Sambucus nigra* Linneo), ruda (*Ruta graveolens* Linneo), sábila (*Aloe vera* Linneo), caléndula (*Calendula officinalis* Linneo), romero (*Rosmarinus officinalis* Linneo), apio (*Apium graveolens* Linneo), manzanilla (*Matricaria chamomilla* Linneo), hierbabuena (*Mentha piperita* Linneo), orégano (*Origanum vulgare* Linneo), menta (*Mentha piperita* Linneo subsp. *citrata* (Ehrh Briq.)), toronjil (*Melissa officinalis* Linneo), mirto (*Psychotria rufescens*), altamisa (*Ambrosia arborescens* Mill.), tomillo (*Pectis graveolens* Klatt), paico (*Chenopodium ambrosioides*), albahaca (*Ocimum basilicum* Linneo), amápola (*Papaver rhoeas* Linneo), prontoalivio (*Lippia alba*-Verbenaceae), cayeno (*Hibiscus rosa-sinensis*) y ají (*Capsicum annuum* Linneo var. *minimum* (Mill.)).

Como complemento, se realizó un intercambio de saberes entre el conocimiento tradicional y el científico con base en la literatura consultada (Alarcón-Restrepo, 2011; Bernal *et al.*, 2011; Castellanos *et al.*, 2017; Duque, 2002; ICA, 2015; Invima, 2019; García *et al.*, 2016; Ministerio de la Protección Social, 2008), lo que permitió la selección de las plantas promisorias.

Por otra parte, se compartió con los participantes el estudio realizado por Córdoba (2017), en el que se presenta un plan de negocios para la creación de la comercializadora de hierbas aromáticas orgánicas "Sabor de Origen". Este trabajo destaca a la albahaca (*Ocimum basilicum* Linneo) (81,25 %), el tomillo (*Pectis graveolens* Klatt) (67,5 %), el laurel (65 %), el orégano (60 %) y el estragón (12,5 %) como las especies de mayor interés para la comercialización a escala internacional, gracias a las preferencias de los consumidores. Como se observa en el listado de plantas de la vereda Altos de Ceylan, algunas de estas especies se cultivan allí.

De otro lado, el 90 % de las personas encuestadas por Córdoba (2017) manifiesta que actualmente compra hierbas aromáticas valorando la frescura del producto, resultado que puede ser aprovechado por los participantes para la selección de las plantas y el proyecto productivo a desarrollar en la vereda. Este hallazgo es respaldado por el Instituto Colombiano Agropecuario (ICA, 2015), que destaca la comercialización y exportación en fresco de plantas medicinales y aromáticas.

Cultivo plantas medicinales

Los integrantes de Asepamuvic trabajaron en tres parcelas cultivando romero (*Rosmarinus officinalis* Linneo), manzanilla (*Matricaria chamomilla* Linneo), hierbabuena (*Mentha piperita* Linneo) y sábila (*Aloe vera* Linneo), actividad para la que adoptaron las BPA recomendadas por la OMS (2003) y además tuvieron en cuenta las condiciones ambientales de la vereda, las características del suelo, la disponibilidad de agua y el tipo de semillas a utilizar. Al tratarse de cultivos ecológicos, se descartó el uso de plaguicidas de origen químico para las labores de manejo agronómico (Amaya *et al.*, 2008).

Análisis multicriterio

Para el cumplimiento de esta actividad se utilizó el listado de plantas medicinales y aromáticas consideradas como relevantes para los propósitos de este proyecto, el cual fue construido de manera conjunta con los participantes. La realización del AMC consideró las categorías ambientales y económicas y los criterios empleados para su valoración, como se detalla a continuación.

Categoría ambiental

Especies nativas

La tabla 2 muestra el detalle de las especies nativas y no nativas identificadas en el sitio estudiado, incluyendo la relación de sus nombres científicos.

Tabla 2. Especies nativas y no nativas

Planta	Nombre científico	Origen	
		Nativa	No nativa
Apio	<i>Apium graveolens</i> Linneo	0	100
Ají	<i>Capsicum annum</i> Linneo var. <i>minimum</i> (Mill.)	0	100
Amapola	<i>Papaver rhoeas</i> Linneo	0	100
Albahaca	<i>Ocimum basilicum</i> Linneo	0	100
Altamisa	<i>Ambrosia arborescens</i> Mill	0	100
Borrachero (Belladona)	<i>Atropa belladonna</i> Linneo	0	100
Caléndula	<i>Calendula officinalis</i> Linneo	0	100

Cayena	<i>Hibiscus rosa-sinensis</i>	0	100
Comfrey	<i>Symphytum officinale</i> Linneo	0	100
Guaba	<i>Phytolacca icosandra</i>	0	100
Hierba del dolor	<i>Acalypha arvensis</i>		
Hierbabuena	<i>Mentha piperita</i> Linneo	0	100
Lengua de suegra milagrosa	<i>Sansevieria trifasciata</i> Prain	0	100
Limonaria o limoncillo	<i>Cymbopogon citratus</i>	0	100
Manzanilla	<i>Matricaria chamomilla</i> Linneo	0	100
Menta	<i>Mentha piperita</i> Linneo subsp. <i>citrata</i> (Ehrh) Briq.	0	100
Mirto	<i>Psychotria rufescens</i>	0	100
Orégano	<i>Origanum vulgare</i> Linneo	0	100
Paico	<i>Chenopodium ambrosioides</i>	100	0
Perejil	<i>Petroselinum crispum</i> (Mill.)	100	0
Poleo	<i>Minthostachys tomentosa</i>	0	100
Prontoalivio	<i>Lippia alba</i> -Verbenaceae	100	0
Ruda	<i>Ruta graveolens</i> Linneo	0	100
Romero	<i>Rosmarinus officinalis</i> Linneo	0	100
Sábila	<i>Aloe vera</i> Linneo	0	100
Sauco	<i>Sambucus nigra</i> Linneo	0	100
Tomillo	<i>Pectis graveolens</i> Klatt	100	0
Toronjil	<i>Melissa officinalis</i> Linneo	0	100

Fuente: elaboración propia con base en información de Bernal *et al.* (2011).

Seguridad del uso de la especie (calidad, efectividad e inocuidad), fuentes, sustentabilidad y calidad de la materia prima

La tabla 3 presenta el estudio bibliográfico realizado a las plantas medicinales. Para ello se tomaron en cuenta los aspectos relacionados con la seguridad del uso de la especie mencionados anteriormente. Además, se revisaron los estudios de bioprospección realizados, el conocimiento tradicional documentado y la sustentabilidad de cada especie, dimensión relacionada con el número de especies promisorias y la manera en que se obtiene la planta medicinal (recolección silvestre o cultivo).

Tabla 3. Seguridad del uso de la especie, aprovechamiento de la biodiversidad y sustentabilidad

Planta	Seguridad del uso de la especie (estudios farmacológicos y toxicológicos)	Seguridad del uso de la especie (inclusión lista Invima)	Estudios de bioprospección	Conocimiento tradicional documentado	Sustentabilidad	Total
Apio	50	50	50	50	10	210
Ají	50	50	50	50	10	210
Amapola	50	0	50	50	2	152
Albahaca	50	50	50	50	10	210
Altamisa	50	0	50	50	2	152
Borrachero (Belladona)	50	50	50	50	5	205
Caléndula	50	50	50	50	10	210
Cayena	50	0	50	50	5	205
Consuelda (Comfrey)	50	0	50	50	5	155
Guaba	50	50	50	50	10	210
Hierba del dolor	50	0	50	50	2	152
Hierbabuena	50	50	50	50	10	210
Lengua de suegra milagrosa	50	0	50	50	5	155
Limonaria o limoncillo	50	50	50	50	10	210
Manzanilla	50	50	50	50	10	210
Menta	50	0	50	50	5	155
Mirto	50	0	50	50	5	155
Orégano	50	50	50	50	10	210
Ortiga	50	50	50	50	10	210
Paico	50	0	50	50	5	155
Perejil	50	50	50	50	10	210
Poleo	50	0	50	50	2	152
Prontoalivio	50	50	50	50	10	210
Ruda	50	50	50	50	5	205
Romero	50	50	50	50	10	210
Sábila	50	50	50	50	10	210
Sauco	50	50	50	50	10	210
Tomillo	50	0	50	50	10	210
Toronjil	50	50	50	50	5	205

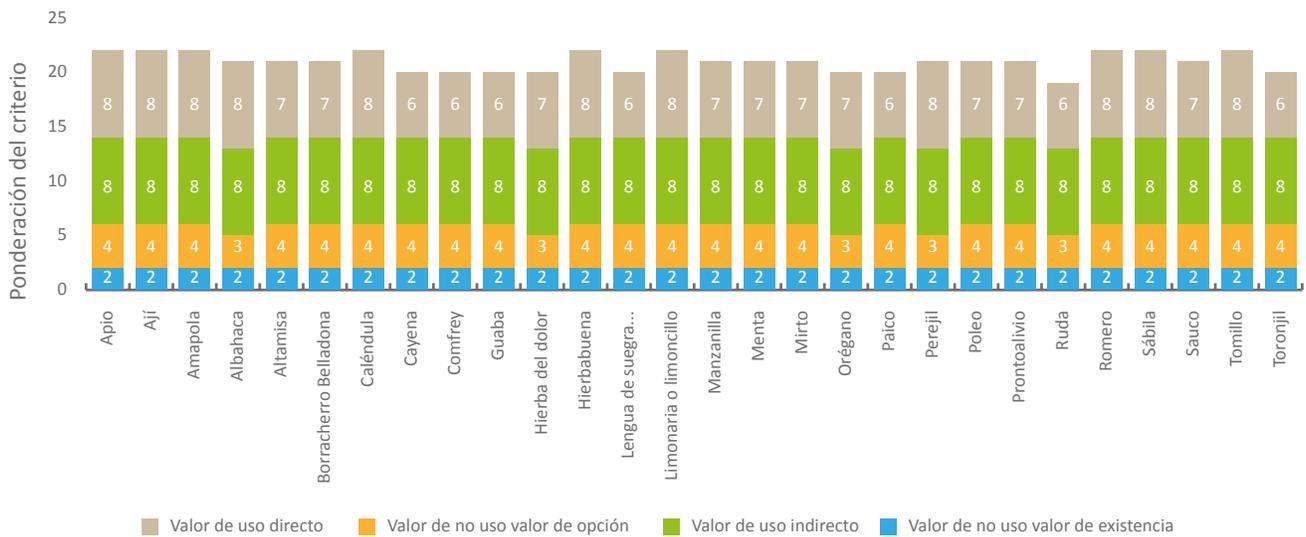
Fuente: elaboración propia.

Categoría económica

Valor económico total (VET)

La utilización del concepto de VET dentro de los criterios de selección se debe a que este permite valorar en su totalidad el aporte de las plantas medicinales al valor económico total de los servicios ambientales de un ecosistema, permitiendo conocer el valor del bien en sí y de sus servicios y funciones como parte de un ecosistema (Barzev, 2002; Camacho & Lancheros, 2008; Pearce & Turner, 1995). El análisis correspondiente se representa en la figura 1.

Figura 1. Valor económico total de las especies estudiadas



Fuente: elaboración propia.

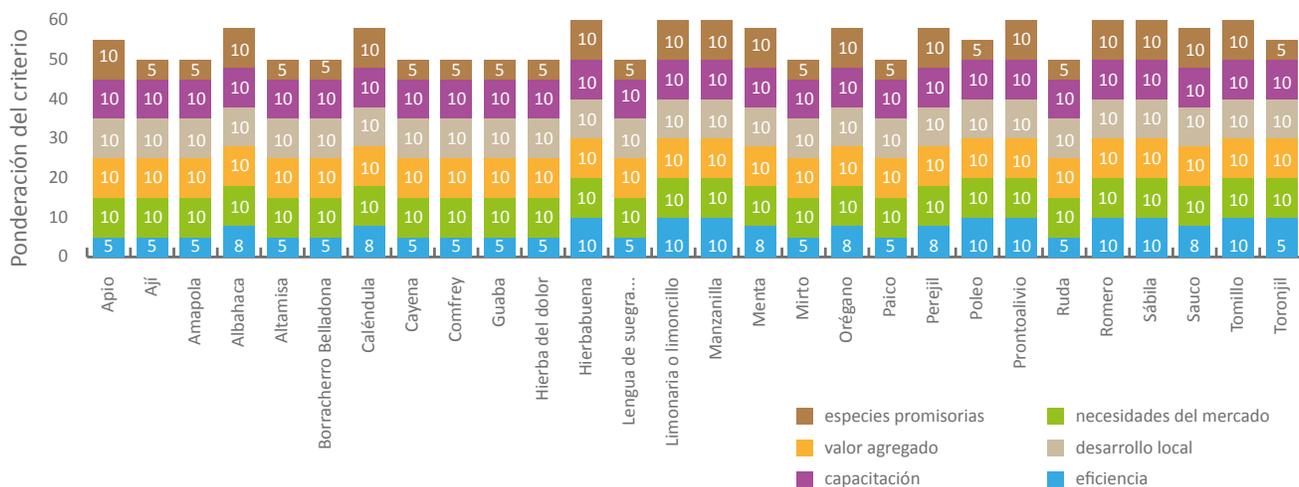
Especies promisorias para uso comercial

Se ha establecido que algunas plantas de especies diferentes pueden tener el mismo nombre común dependiendo de las ciudades donde se comercialicen. Sin embargo, no se cuenta con información precisa sobre el volumen de las diferentes especies de plantas consumidas o comercializadas (Díaz, 2003).

El criterio ambiental toma factores como la seguridad de uso de la planta, el aprovechamiento de la biodiversidad y la fuente, y la sustentabilidad y calidad de la materia prima, lo cual nos orientó a la selección de las plantas medicinales sábila, limoncillo, prontoalivio, manzanilla, hierbabuena, tomillo, perejil y romero.

En el criterio económico se tiene como referencia el valor de uso y el valor de no uso, seleccionando las mismas plantas mencionadas en el criterio ambiental debido a su aporte significativo al valor de no uso (valor de opción y valor por existencia), producto de su potencial farmacéutico e industrial. Desde el valor de uso, estas especies cuentan con diversos usos indirectos que dan soporte al ecosistema y, así mismo, contribuyen al valor de uso directo como plantas medicinales, materias primas, alimentos y plantas ornamentales. Además, estas plantas no corresponden a cultivos extensivos, poseen aplicaciones generalmente específicas e incluyen un aporte considerable de valor agregado en conocimiento tradicional y mano de obra, lo cual se enriquece con procesos de investigación, bioprospección y capacitación desarrollados para cultivos ecológicos, aportando así al desarrollo local, regional y nacional a través de la educación y la gestión ambiental (García *et al.*, 2016), como se muestra en la figura 2.

Figura 2. Categorías económicas



Fuente: elaboración propia.

DISCUSIÓN

El análisis multicriterio aplicado a la ponderación de los criterios ambientales y económicos necesarios para seleccionar un grupo de plantas de interés a ser involucradas en procesos productivos y programas de certificación ecológica, teniendo en cuenta la política Nacional de Biodiversidad y sus principios (conocer, conservar, restaurar y utilizar) (MADS, 2012), acudió a la apropiación social del conocimiento a través del diálogo de saberes. En lo ambiental, se dio prioridad al origen nativo de la especie seleccionada para aprovechar la biodiversidad de Colombia en términos de plantas medicinales, así como la bioprospección y el conocimiento tradicional como valor agregado, sumados al concepto de valor económico total (VET) (Pearce & Turner, 1995).

Las plantas medicinales seleccionadas son recolectadas en su mayoría en forma silvestre, sin control alguno o la realización de capacitaciones para los recolectores, quienes en su mayoría se han dedicado a esta labor por años. Durante el proceso se evidencia la falta de cumplimiento de pautas elementales como recolectar las plantas en el momento adecuado de su desarrollo vegetal o no llevar a cabo esta labor cerca de carreteras o zonas industriales. También se da cuenta de un número importante de personas involucradas en el cultivo de estas especies que no tiene

experiencia o capacitación alguna, lo que influye en el impacto ecológico de esta actividad agrícola y a su vez difunde un manejo inadecuado para la obtención de plantas de uso medicinal de calidad. Esto resulta importante debido a que en los últimos años la demanda de plantas medicinales se ha multiplicado y la población rural o citadina acude a este tipo de actividades para subsistir (Kate & Laird, 2019).

Como parte del trabajo desarrollado en la vereda Altos de Ceylan se construyó una matriz aprender-desaprender sobre plantas medicinales y aromáticas, similar a lo propuesto por autores como Gallegos-Zurita (2016), quien considera a las plantas medicinales como la principal alternativa para el cuidado de la salud. Además, esta investigación permitió conocer métodos de cultivo basados en el conocimiento tradicional, la aplicación de BPA (OMS, 2003), el manejo de la gestión ambiental y el control de plaguicidas químicos (Amaya *et al.*, 2008) mediante el uso alternativo de bioplaguicidas (Camacho *et al.*, 2010).

Los participantes del estudio construyeron herbarios y apreciaron su propio saber al evidenciar que este era valorado y que sus opiniones eran tomadas en cuenta para la toma de decisiones. Estas estrategias han sido desplegadas por autores como Sandoval (2018), quien muestra la importancia de los procesos educativos participativos y horizontales basados en saberes

y prácticas locales en asociaciones campesinas como Agruco, donde se resalta la agricultura biológica, sana y limpia para la conservación del medio ambiente y el fortalecimiento de las bases productivas, mejorando la calidad de vida de las familias campesinas. El estudio realizado con Asepamuvic hizo posible romper asimetrías, especialmente aquellas entre los saberes campesinos y los saberes técnicos, puesto que ambas se lograron fusionar para la construcción de miradas colectivas integrales, como se aborda en estudios similares (García-de-Alba *et al.*, 2012; Leff *et al.*, 2004; Rivera *et al.*, 2017; Sandoval, 2018).

Se enunciaron las especies de plantas medicinales identificadas tal y como se denominan localmente, complementando esta información con el saber científico dado por el Invima (2019), el Instituto de Investigación de Recursos Biológicos Alexander von Humboldt (Bernal *et al.*, 2011) y el ICA (Alarcón-Restrepo, 2011; ICA, 2015). Por lo anterior, se recomienda realizar una identificación certificada de estas especies a fin de llevar su trazabilidad y poder establecer aquellas que se encuentran y se cultivan en la vereda, como fue hecho por Vera y Sánchez (2015), quienes registraron algunas plantas medicinales cultivadas en el corregimiento de San Cristóbal, municipio de Medellín.

Con esta selección se pretende mostrar el potencial de la biodiversidad de nuestro país con respecto a las variedades comercializadas en otros países y llamar la atención hacia variedades nativas aún no explotadas, estimulando su cultivo y permitiendo cumplir la posible demanda de especies promisorias como el prontoalivio (*Lippia alba*-Verbenaceae) y el perejil (*Petroselinum crispum* (Mill.)).

Para esta investigación se diseñaron talleres de capacitación con los miembros de la asociación Asepamuvic orientados a establecer la estrategia pedagógica, la metodología de trabajo, las técnicas y los contenidos a abordar. De esta manera, se generaron diversos materiales impresos que contribuyeron a que los participantes pudieran profundizar en el abordaje de las distintas temáticas tratadas. Esto concuerda con lo trabajado por Rivera (2017), quien brindó espacios de diálogo donde los participantes compartieron sus problemas, experiencias y expectativas sobre el manejo de la vida silvestre, aprendiendo con ello sobre el manejo técnico de esta actividad.

Las plantas medicinales tienen como beneficios monetarios el incremento de precios, avances en pagos para la realización de estudios de investigación, pagos por propiedad intelectual y otros beneficios financieros (investigación + desarrollo). Como beneficios no monetarios se resalta la provisión de infraestructura, equipos, tecnologías y asistencia a programas enfocados a estandarizar los sistemas médicos tradicionales, así como el mejoramiento del manejo de especies nativas. Actualmente, el mercado de este tipo de productos recibe apoyo institucional para el proceso de certificación ecológica por parte de la Federación Nacional de Cafeteros, Cenicafé, la Corporación Colombia Internacional, el Consejo de Administración Forestal (FSC), Biointropic, el Departamento de Agricultura de los Estados Unidos (USDA) y EurepGAP. Por ello, se espera lograr dicha certificación e incrementar la competitividad de estos productos tanto en el entorno nacional como en los mercados internacionales (MADR, 2019).

En cuanto a los participantes del proyecto, se rescata que los tiempos de las comunidades están basados en ciclos naturales y políticos relacionados con la dinámica de sus comunidades, donde la prioridad es el futuro inmediato de sus familias, la productividad de sus actividades y el beneficio de la comunidad. Estos factores fueron mencionados por los participantes como motivaciones para continuar haciendo parte de este tipo de iniciativas, lo cual es particularmente relevante al considerar la participación de las mujeres rurales, que en este trabajo fue bastante destacada.

CONCLUSIONES

El uso del análisis multicriterio y el diálogo de saberes permitió enriquecer el trabajo realizado con Asepamuvic frente a la utilidad y la valoración ambiental y económica de las plantas aromáticas y medicinales presentes y cultivadas en la vereda Altos de Ceylan, ubicada en el municipio de Viotá (Cundinamarca, Colombia), como parte de un proceso productivo.

De acuerdo con el AMC efectuado, las plantas seleccionadas para la siembra de cultivos y su inclusión en programas de certificación ecológica son sábila (*Aloe vera* Linneo), limoncillo (*Cymbopogon citratus*), prontoalivio (*Lippia alba*-Verbenaceae), manzanilla (*Matricaria chamomilla* Linneo), hierbabuena (*Mentha piperita* Linneo), tomillo (*Pectis graveolens*

Klatt), perejil (*Petroselinum crispum* (Mill.)) y romero (*Rosmarinus officinalis* Linneo). Dentro de este grupo se destacan el prontoalivio y el perejil, plantas nativas que se encuentran en el listado del Invima (2019).

AGRADECIMIENTOS

Las autoras agradecen a los miembros de Asepamuvic por su participación entusiasta en el desarrollo de este trabajo de investigación. También agradecen a la Universidad Colegio Mayor de Cundinamarca por el apoyo recibido durante su realización.

REFERENCIAS

- Alarcón-Restrepo, J. J. (2011). *Plantas aromáticas y medicinales: enfermedades de importancia y usos terapéuticos. Medidas para la temporada invernal*. ICA.
- Amaya, E. F., Roa, A. M., Camacho, J. E., & Meneses, S. (2008). Valoración de factores de riesgo asociados a los hábitos de manejo y exposición a organofosforados y carbamatos en habitantes y trabajadores de la vereda de Bateas del municipio de Tibacuy, Cundinamarca, Colombia. *Nova*, 6(10), 147-155. <https://doi.org/10.22490/24629448.405>
- Barzev, R. E. (2002). *Guía metodológica de valoración económica de bienes, servicios e impactos ambientales. Un aporte para la gestión de ecosistemas y recursos naturales en el CBM*. Corredor Biológico Mesoamericano.
- BBVA (2017, agosto 8). *Descubre las tecnologías sociales para entornos colaborativos*. <https://www.bbva.com/es/descubre-tecnologias-sociales-entornos-colaborativos/>
- Bernal, H. Y., García-Martínez, H., & Quevedo-Sánchez, G. F. (2011). *Pautas para el conocimiento, conservación y uso sostenible de las plantas medicinales nativas en Colombia: Estrategia nacional para la conservación de plantas*. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt.
- Camacho, J. E., & Castiblanco, C. (2008). Certificación ecológica de dos plantas medicinales como un paso hacia el aprovechamiento de la biodiversidad en Colombia. *Medio Ambiente y Desarrollo. Aporte de los estudiantes del PMAD al debate ambiental*. Universidad Nacional de Colombia.
- Camacho, J. E., Gómez, M. I., & Villamizar, L. F. (2010). Selección de un sistema de atomización para la formación de micropartículas de eudragit® s100 en lecho fluido. *Nova*, 8(13), 87-100. <https://doi.org/10.22490/24629448.442>
- Castellanos-Castro, C., Sofrony, C., & Higuera, D. (eds.). (2017). *Plan de acción de la Estrategia Nacional para la Conservación de Plantas de Colombia*. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Ministerio de Ambiente y Desarrollo Sostenible, & Red Nacional de Jardines Botánicos de Colombia. <https://doi.org/10.21068/B001.2016.304>
- Córdoba, R. (2017). *Plan de negocio para la creación de la comercializadora de hierbas aromáticas orgánicas - Sabor de Origen* (tesis de maestría, Universidad Santo Tomás). Repositorio USTA. <https://repository.usta.edu.co/handle/11634/9784>
- Delgado, F., & Rist, S. (eds.). (2016). *Ciencias, diálogo de saberes y transdisciplinariedad*. Agroecología Universidad Cochabamba.
- Díaz, J. (2003). *Informe Técnico. Caracterización del Mercado Colombiano de Plantas Medicinales y Aromáticas. I*. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt.
- Duque, A. (2002). Informe del primer seminario iberoamericano de comercialización de plantas medicinales y aromáticas. *Boletín Latinoamericano y del Caribe de plantas Medicinales y Aromáticas*, 1(2), 12-14. <https://www.redalyc.org/pdf/856/85610205.pdf>
- Gallegos-Zurita, M. (2016). Las plantas medicinales: principal alternativa para el cuidado de la salud, en la población rural de Babahoyo, Ecuador. *Anales de la Facultad de Medicina*, 77(4), 327-332. <https://doi.org/10.15381/anales.v77i4.12647>

- García-de-Alba, J. E., Ramírez-Hernández, B. C., Robles-Arellano, G., Zañudo-Hernández, J., Salcedo-Rocha, A. L., & García-de-Alba-Verduzco, J. E. (2012). Conocimiento y uso de las plantas medicinales en la zona metropolitana de Guadalajara. *Desacatos*, 39, 29-44. <https://doi.org/10.29340/39.238>
- García-Martínez, H., Moreno, L. A., Londoño, C., & Sofrony, C. (2016). *Estrategia nacional para la conservación de plantas: actualización de los antecedentes normativos y políticos y revisión de avances*. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt.
- Hernández-Rincón, E. H., Lamus-Lemus, F., Carratalá-Munuera, C., & Orozco-Beltrán, D. (2017). Diálogo de saberes: propuesta para identificar, comprender y abordar temas críticos de la salud de la población. *Revista Salud Uninorte*, 33(2), 242-251. <https://doi.org/10.14482/sun.33.2.10552>
- Instituto Colombiano Agropecuario [ICA]. (2015, octubre 20). *Predios productores de frutales, hortalizas y hierbas aromáticas registrados en el ICA para exportación en fresco*. <https://www.ica.gov.co/icacomunica/prensa/el-registro-de-predios-productores-agropecuarios-f>
- Instituto Nacional de Vigilancia de Medicamentos y Alimentos [Invima]. (2018). *Listado de plantas medicinales aceptadas con fines terapéuticos*. <https://www.invima.gov.co/documents/20143/461703/LISTADO-DE-PLANTAS-JULIO-31-2018.pdf/7bdad908-d5d8-76aa-c2bf-41590611066a>
- Jiménez-Herrero, L. M. (1997). *Desarrollo sostenible y economía ecológica*. Síntesis.
- Kate, K., & Laird, S. A. (2019). *The commercial use of biodiversity: Access to genetic resources and benefit-sharing*. Routledge. <https://doi.org/10.4324/9780429341540>
- La República (2018) *La producción de los cultivos de plantas aromáticas y especias crecieron 21% en 2017*. <https://www.larepublica.co/economia/laproduccion-de-los-cultivos-de-plantas-aromaticas-y-especias-crecieron-21-en2017-2771263>
- Leff, E. (2004). Racionalidad ambiental y diálogo de saberes. Significancia y sentido en la construcción de un futuro sustentable. *Polis. Revista Latinoamericana*, 2(7), 13-40. <https://doi.org/10.5380/dma.v7i0.3042>
- Ministerio de Agricultura y Desarrollo Rural [MADR]. (2019). *Lista de organismos de certificación acreditados y autorizados para certificar bajo el reglamento para la producción ecológica adoptado por la Resolución 0187/2006, y autorizados para administrar el uso del Sello de Alimento Ecológico*. MADR https://www.minagricultura.gov.co/tramites-servicios/Documents/Lista_de_Certificadoras_de_Productos_Ecologicos_040219.pdf
- Ministerio de Agricultura y Desarrollo Rural [MADR]. (2019). *Cadena de plantas aromáticas, medicinales, condimentarias y afines. Dirección de cadenas agrícolas y forestales*. MADR.
- Ministerio de Ambiente y Desarrollo Sostenible [MADS]. (2012). *Política Nacional para la Gestión Integral de la Biodiversidad y sus Servicios Ecosistémicos*. MADS.
- Ministerio de la Protección Social de Colombia (2008). *Vandemécun Colombiano de plantas medicinales*. Imprenta Nacional de Colombia.
- Morse, J. M. (1994). Emerging from the data: The cognitive processes of analysis in qualitative inquiry. En J. Morse (ed.), *Critical issues in qualitative research methods*, (pp. 23-46). Sage. <https://doi.org/10.2307/2076454>
- Ochoa, D. A. R. (2012). La salud pública como escenario para el diálogo de saberes. *Revista CES Salud Pública*, 3(1), 1-3.
- Organización Mundial de la Salud [OMS]. (1988). *Declaración de Chiang Mai, "Salvar vidas salvando las plantas"*. OMS.
- Organización Mundial de la Salud [OMS]. (2003). *Directrices de la OMS sobre buenas prácticas agrícolas y de recolección (BPAR) de plantas medicinales*. OMS.

- Pearce, D. W., & Turner, P. K. (1995). *Economía de los recursos naturales y del medioambiente*. Celeste.
- Programa de las Naciones Unidas para el Desarrollo [PNUD]. (2011). *Mujeres rurales: gestoras de esperanza*. PNUD
- Reyes-Sánchez, G. L. (ed.). (2009). *Diálogo de saberes: plantas medicinales, salud y cosmovisiones*. Universidad Nacional de Colombia.
- Rincón-Parra, N. S. (2012). Contexto y perspectiva de la red de suministro: Plantas aromáticas en Colombia. *ECONÓMICAS CUC*, 33(1), 135-156.
- Rivera, A. P. T., Ortega-Cuadros, M., Melo-Ríos, A., & Mier-Giraldo, H. J. (2017). Vigilancia tecnológica de plantas aromáticas: de la investigación a la consolidación de la agrocadena colombiana. *Ciencia & Tecnología Agropecuaria*, 18(2), 353-377. https://doi.org/10.21930/rcta.vol18_num2_art:636
- Romero, M. J., & Rodríguez, E. B. (2010). Diálogo de saberes: Medicina tradicional y medicina occidental moderna. *Revista de Investigaciones UNAD*, 9(2), 125-133. <https://doi.org/10.22490/25391887.676>
- Sandoval, R. J. C. (2018). Gestión intercultural de la biodiversidad: investigación acción en un proceso educativo con campesinos de Veracruz. *Revista Interamericana de Educación de Adultos*, 40(2), 119-135. https://www.redalyc.org/pdf/4575/Resumenes/Resumen_457556293006_1.pdf
- Sipahi, S., & Mehpare, T. (2010). The analytic hierarchy process and analytic network process: an overview of applications. *Management Decision*, 48(5), 775-808. <https://doi.org/10.1108/00251741011043920>
- Vera, M., & Sánchez, S. M. (2015). Registro de algunas plantas medicinales cultivadas en San Cristóbal, municipio de Medellín (Antioquia, Colombia). *Revista Facultad Nacional de Agronomía*, 68(2), 7647-7658. <https://doi.org/10.15446/rfnam.v68n2.50979>

¿Ganadería o conservación? Caracterización del conflicto ambiental en la Reserva Forestal Protectora “El Malmo” (Tunja, Boyacá, Colombia)

Livestock or Conservation? Characterization of the Environmental Conflict in the Protective Forest Reserve “El Malmo” (Tunja, Boyacá, Colombia)

Albaluz Ramos Franco^{ab}, Inti Alba-Vargas^{ac}, Leyde Cabrera-Ruiz^{ad}
y Camila Díaz-Rojas^e

^a Grupo de Investigación Biología para la Conservación, Facultad de Ciencias Básicas, Universidad Pedagógica y Tecnológica de Colombia, Tunja, Colombia

^b <http://orcid.org/0000-0001-6432-4061> | albaluz.ramos@uptc.edu.co

^c <https://orcid.org/0000-0002-5569-2214> | inti.alba@uptc.edu.co

^d <https://orcid.org/0000-0003-2223-6877> | leyde.cabrera@uptc.edu.co

^e Grupo de Investigación Unidad de Ecología en Sistemas Acuáticos (UDESA), Facultad de Ciencias Básicas, Universidad Pedagógica y Tecnológica de Colombia, Tunja, Colombia | <https://orcid.org/0000-0001-9134-7453> | camila.diaz@uptc.edu.co

RESUMEN

La Reserva Forestal Protectora El Malmo es fuente de pugna entre las autoridades ambientales y los habitantes de la vereda Barón Germania (Tunja, Boyacá, Colombia), pues esta encuentra en un paisaje rural dominado por actividades agropecuarias que están en contravía con el objetivo de conservación del área protegida. El Plan de Manejo de la Reserva, tras cuatro décadas de declaración, no ha sido adoptado y la Corporación Autónoma Regional de Boyacá no ha definido las zonas donde las actividades económicas rurales están prohibidas, generando desconfianza y temor, en razón de las acciones correctivas y policivas por parte de la autoridad ambiental. En esta investigación se

realiza una caracterización del conflicto ambiental en el área de estudio que expone actores y dimensiones. Además, se hace énfasis en la ausencia de mecanismos de participación ciudadana y de acciones de resolución mediante el desarrollo conjunto de la zonificación en cuanto a restricciones de uso de los recursos naturales existentes en la reserva.

Citation: Ramos-Franco, A., Alba-Vargas, I., Cabrera-Ruiz, L. y Díaz-Rojas, C. (2021). ¿Ganadería o conservación? Caracterización del conflicto ambiental en la Reserva Forestal Protectora “El Malmo” (Tunja-Boyacá). *Mutis*, 11(1), 37-47. <https://doi.org/10.21789/22561498.1638>

Recibido: Agosto 6, 2020. **Aceptado:** Enero 15, 2021.

Copyright: ©2021. Chávez-Plazas, Y. A., Ramírez-Mahecha, M. C. y Camacho-Kurmen, J. E. (2021). This is an open-access article, which permits unrestricted use, distributions and reproduction in any medium, provided the original author and source are credited.

Competing Interests: The authors have no conflict of interest.

Palabras clave: conflicto ambiental, Reserva Forestal Protectora El Malmo, delimitación y zonificación, participación ciudadana, uso del suelo.



ABSTRACT

“El Malmo” Protective Forest Reserve is a source of conflict between environmental authorities and the inhabitants of the Barón Germania settlement (Tunja, Boyacá, Colombia), as it is located in a rural landscape ruled by agricultural activities that go against the conservation purposes of the protected area. Although a reserve management plan for this area was enacted four decades ago, it has not been adopted to this date. Besides, Boyacá regional environmental authority (Corpoboyacá) has not defined the areas where rural economic activities are prohibited, generating distrust and fear, due to the corrective actions deployed. For these reasons, this research presents a characterization of the environmental conflict in the study area that exposes actors and dimensions, emphasizing the absence of citizen participation mechanisms and resolution actions by jointly developing the zoning in terms of restrictions to the use of existing natural resources in the reserve.

Keywords: Environmental conflict, El Malmo Protective Forest Reserve, delimitation and zoning, citizen participation, land use.

INTRODUCCIÓN

La ganadería está ligada a la historia de Colombia pese a que ha disminuido su participación en la economía nacional frente a otros sectores productivos (Cuenca-Jiménez *et al.*, 2008; Murgueitio, 2003; Vergara-Vergara, 2010). Actualmente, esta actividad ocupa un área aproximada de 43 millones de hectáreas en el país (DANE, 2016). En la región Andina, la producción ganadera es de tipo intensiva y se ha desarrollado en tierras aptas para usos agrícolas y forestales, lo cual evidencia la subutilización de recursos, no solo del suelo, sino también del agua y del aire (Mora-Marín *et al.*, 2017).

Entre los impactos atribuidos a esta actividad se destaca la degradación del suelo, definida como la disminución de la capacidad de producción o de su calidad (Mora-Marín *et al.*, 2017; Murgueitio, 1999, 2003; Murgueitio & Ibrahim, 2008). Existe una división de los disturbios sobre el suelo, bien sean estos directos (animal por pisoteo o excreciones) o indirectos, en razón del aporte del hombre sobre el terreno (Doran & Parkin, 1994; Lal *et al.*, 2012), señalando que la ganadería contribuye a la transformación de los ecosistemas naturales y que esta guarda una fuerte relación con la tala y quema de árboles (Mora-Marín *et al.*, 2017; Murgueitio, 2003).

La preocupación por conservar relictos de ecosistemas con el objeto de garantizar la oferta de sus servicios ecosistémicos por más tiempo ha incrementado en los últimos años (MinAmbiente, 2012); esto mediante acciones de manejo, distribución y administración de los recursos naturales por parte de diferentes

entes gubernamentales, a través de actos administrativos como la declaratoria de áreas protegidas estatales. Estas áreas se entienden como “aquella zona delimitada geográficamente designada o regulada y administrada con el fin de alcanzar objetivos específicos de conservación” (Álvarez, 2011, p. 27). En relación con las reservas forestales, en los artículos 205 y siguientes del Código de Recursos Naturales Renovables (Decreto Ley 2811 de 1974), se les define como “aquellas zonas de propiedad pública o privada destinadas para el establecimiento o mantenimiento y utilización racional de áreas forestales productoras o protectoras”.

Aunque las áreas protegidas solucionan en parte el problema de la conservación, su declaratoria genera nuevos conflictos ambientales que no solo tienen implicaciones ecológicas, sino también sociales, económicas y culturales (Rodríguez, 2016). Desde el momento de su declaratoria, los límites de las reservas pueden superponerse con predios, generalmente habitados por comunidades campesinas, con los cuales el Estado no ha podido negociar su compra u otra forma de compensación a cambio de las restricciones de uso del suelo que traen consigo. La Reserva Forestal Protectora “El Malmo” (Tunja, Boyacá, Colombia) ilustra lo anterior, al ser escenario de un conflicto ambiental a nivel local dada su ubicación en un área dominada por actividades ganaderas y agrícolas de tradición, las cuales se convierten en una amenaza para los fines de conservación previstos.

Este trabajo tiene como objetivo caracterizar el conflicto ambiental de la ganadería frente a la conservación en la Reserva Protectora “El Malmo” a través de

la tipificación del conflicto ambiental, la identificación de los actores implicados y el reconocimiento de las dimensiones temporal, biótica, social, económica y legal que le conforman. Adicionalmente, se busca proponer posibles alternativas de solución a este conflicto socioambiental.

METODOLOGÍA

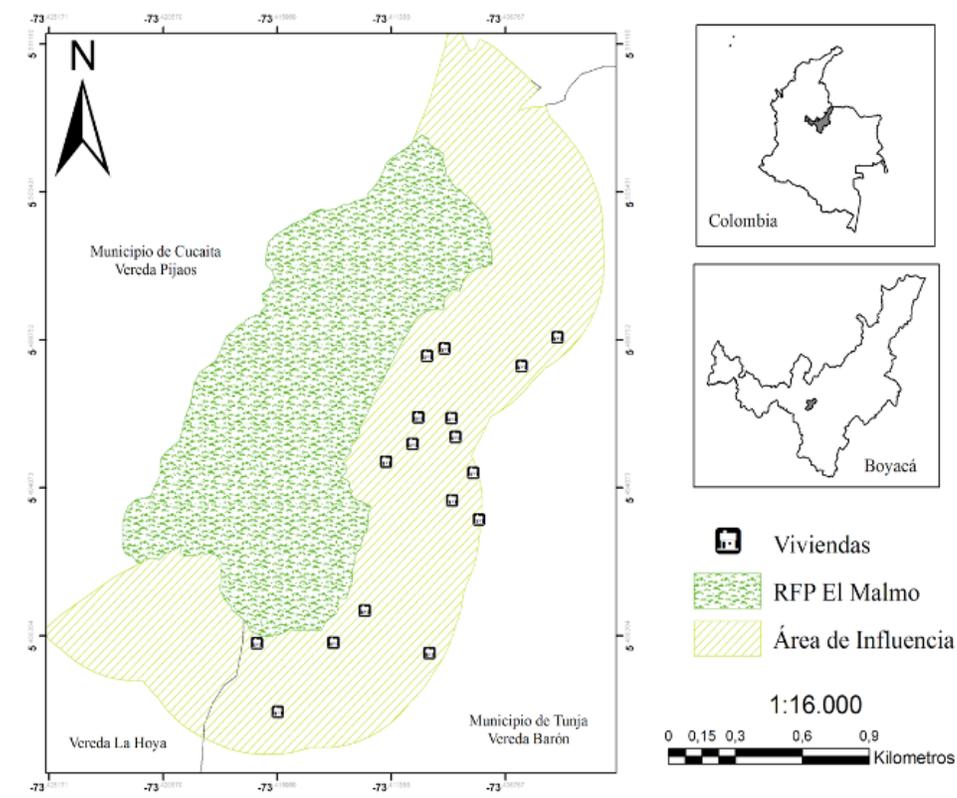
Área de estudio

La Reserva Forestal Protectora “El Malmo” se localiza a 8 kilómetros del casco urbano de Tunja (Boyacá), por la vía Tunja-Bogotá, en la vereda Barón Germania (05°29’59” N y 73°25’00” W), a una altura entre 3.050 y 3.200 m s. n. m. sobre la falda suroccidental de la cuchilla conocida como “Alto del Muerto”. Esta reserva tiene una vegetación compuesta por cuatro tipos de coberturas: bosque denso altoandino, chuscales, plantaciones de acacias y pastizales (Ramos & Armenteras, 2019).

La definición del área de influencia se realizó con ayuda del software ArcGis 10.5 (ESRI, 2015), aplicando un *buffer* de 500 m de ancho en torno al límite de la reserva y sin salir del territorio del municipio de Tunja. Dentro de dicho *buffer*, y con ayuda de imágenes satelitales, se ubicaron los puntos que representaban edificaciones, enumerándolas de 0 a 16 para su posterior identificación (figura 1). Estos puntos fueron ingresados al GPS para lograr precisión de ubicación durante la fase de trabajo de campo.

La metodología utilizada en este trabajo fue de carácter cualitativo, mediante las técnicas de revisión documental, entrevistas y sesiones colectivas. La investigación está dividida en dos fases: revisión de información secundaria e inmersión social en campo.

Figura 1. Ubicación del área de estudio



Fuente: elaboración propia.

Revisión de información secundaria

En la fase *ex situ* se realizó una revisión normativa, donde se consultó el tipo de reserva a la que corresponde El Malmo, cuándo y cuáles entidades la declararon, así como las condiciones que rigen el manejo y la administración de la zona. Adicionalmente, se registró el estado del proceso legal del Plan de Manejo Ambiental (PMA). Por otra parte, se efectuó un análisis exhaustivo del “Documento técnico actualización del Plan de Manejo Reserva Forestal del Malmo de acuerdo con lo establecido en el Decreto 1076 de 2015” (AGS, 2018).

El marco conceptual utilizado en esta investigación está basado en lo propuesto por Rodríguez (2016), quien define un conflicto ambiental como “una confrontación social, económica y política entre diferentes actores, que surge por la existencia de diversos intereses relacionados con el uso, manejo, aprovechamiento, exploración, explotación, conservación, protección, administración o afectación de los recursos naturales y el ambiente” (p. 10). Este autor también indica que para conocer y plantear una solución a esta problemática se requiere identificar los actores involucrados, la tipología del conflicto y las dimensiones, modificadas hacia un marco temporal, biótico, social, económico y legal.

Fase de campo

El primer encuentro con la comunidad surgió cuando la Corporación Autónoma Regional de Boyacá (Corpoboyacá) realizó una reunión informativa sobre el diagnóstico preliminar del PMA de la Reserva. Durante este encuentro se tomó nota de las reacciones y posición de cada una de las partes presentes frente a cada aspecto incluido dentro del estudio que realizó esta autoridad (figura 2). Posteriormente, con el fin de lograr un acercamiento a la percepción local e incorporar algunos datos más específicos, se realizaron dos visitas a la zona de estudio para aplicar entrevistas semiestructuradas (con su respectivo consentimiento informado) a la comunidad que habita en el área de influencia de la reserva. Además de las respuestas abiertas, se tuvieron en cuenta los comentarios descriptivos de los entrevistados.

Las preguntas realizadas a los individuos cabeza de hogar entrevistados fueron:

- ¿Hace cuánto vive usted en la reserva?
- ¿Es usted ganadero(a)? ¿Desde hace cuánto? Si es así, ¿cuántas cabezas de ganado tiene? En términos monetarios, ¿cuánto representan?
- ¿Sabe usted qué es un área de reserva forestal protectora? ¿Desde hace cuánto conoce de la existencia de la RFP “El Malmo”?
- ¿Se ha visto en la necesidad de ingresar su ganado a la reserva?
- ¿Conoce usted algún tipo de sanción que acarree usar el suelo de una reserva sin permiso? ¿Cuál(es)?
- ¿Qué autoridades o entidades han hecho presencia en la zona? ¿De qué forma?
- ¿Ha recibido alguna capacitación/taller/reunión por parte de las autoridades ambientales?
- ¿Pertenece usted a algún tipo de asociación comunitaria? (acueducto, veredal, comunal, de seguridad, etc.) ¿Sabe de alguna iniciativa que las asociaciones del sector hayan efectuado para promover el cuidado de la Reserva?
- ¿Estaría dispuesto(a) a aportar para el financiamiento de estrategias para la conservación de la Reserva?

Figura 2. Fotografías de la reunión informativa sobre el diagnóstico preliminar del PMA por parte de Corpoboyacá



Fuente: elaboración propia.

RESULTADOS

Inmersión social

De acuerdo con los puntos establecidos en el mapa (figura 1), se identificaron 17 edificaciones, de las cuales 11 corresponden a caballerizas, bodegas y un punto de potabilización de agua. En las construcciones habitadas se recolectó un total de ocho entrevistas. Durante estos acercamientos se indagó, como primera medida, sobre la situación socioeconómica de la persona cabeza de hogar, con preguntas relacionadas con variables como género, edad, nivel educativo, empleo y acceso a servicios de salud.

De otro lado, se obtuvo que el 75 % de los encuestados tienen como actividad productiva principal la ganadería, quienes la practican desde hace bastante tiempo, en rangos desde 10 a 65 años. A su vez, la cantidad de semovientes varía de 6 hasta 90 bovinos, con una representación monetaria entre 5 y 25 millones de pesos colombianos. Del total de entrevistados, 25 % admite haber ingresado los animales de su propiedad a la zona de conservación (siendo uno de ellos el poseedor del mayor número de semovientes en la zona) (figura 3). Por su parte, 52 % de los encuestados desconoce las sanciones que puede acarrear el uso indebido de los predios de la Reserva. Al preguntarles acerca de la definición de una reserva forestal protectora solo 12 % de los encuestados manifestaron saberlo; de estos, el 62 % conoce de la existencia de una zona de especial protección denominada “El Malmo” desde su edad escolar.

Figura 3. Presencia de semovientes en zona de pasturas dentro de la reserva



Fuente: elaboración propia.

La cooperación de la comunidad bajo figuras asociativas es relativamente baja, solo un 25 % de los entrevistados pertenece a alguna de estas organizaciones, mencionando únicamente la Junta de Acción Comunal de la vereda y una asociación productora de leche, por lo que en su mayoría ignoran si se han llevado a cabo proyectos relacionados con el cuidado del área protegida. Finalmente, al preguntar por la disposición a pagar el financiamiento de estrategias para la conservación de la Reserva, la respuesta fue negativa, pues seis de los ocho entrevistados pagarían menos de la mitad de un salario mínimo diario legal vigente (COP 781.242).

Revisión documental

La declaratoria de la Reserva fue realizada por el Instituto Nacional de Recursos Naturales Renovables y del Ambiente (INDERENA), mediante el Acuerdo 036 del 28 de octubre de 1976, posteriormente aprobado por la Resolución 362 del 17 de diciembre de 1976 del Ministerio de Agricultura, bajo el carácter de Reserva Forestal Protectora Nacional (Álvarez, 2011). Actualmente, su administración está bajo la jurisdicción de Corpoboyacá.

Mediante el “Documento técnico de actualización del Plan de Manejo Reserva Forestal del Malmo de acuerdo a lo establecido en el Decreto 1076 de 2015”, Corpoboyacá presenta la información geográfica, geológica, biológica y social de la Reserva, que según esta autoridad sería la base para la ordenación y zonificación de El Malmo. Previo a su publicación, se llevó a cabo una socialización del contenido del documento ante la comunidad de la vereda Barón Germania, con la participación de entes como la Alcaldía de Tunja, la Policía Metropolitana de Tunja, asociaciones comunitarias y otras partes interesadas. En este encuentro además se mencionaron dos metas importantes a partir del PMA: la primera aborda la protección de la Reserva mediante la compra de predios cercanos al área protegida y actividades de reforestación en los mismos; la segunda se dirige a incorporar la reconversión agropecuaria, la cual debe ser concertada con la comunidad. Esta última fue un punto de inflexión que generó muchas inquietudes en la comunidad, puesto que las personas manifestaron desconocimiento con respecto a la delimitación e imposición de zonas restrictivas de actividad agropecuaria dentro de la reserva. Por su parte, los campesinos manifestaron

su inconformismo, denunciando casos cercanos de desalojo de predios tras la delimitación del páramo de Rabanal (Resolución 1768 de 2016) por parte del Ministerio de Ambiente, ya que esta declaratoria no les permite continuar con su actividad de subsistencia, pero tampoco les proporciona los medios alternativos de economía familiar que se prometieron en los estudios técnicos. Es así como la reunión se vio interrumpida por el disgusto de los asistentes.

Caracterización del conflicto ambiental

La tipología de conflicto ambiental, según su raíz —y siguiendo la metodología de análisis propuesta por Rodríguez (2016)—, corresponde a “declaración o sustracción de áreas protegidas” y “decisión de las autoridades”, puesto que la expedición del acto administrativo mediante el cual se declara la Reserva impone nuevas reglas de uso y no uso sobre el suelo, lo cual se plasmará una vez se adopte el correspondiente PMA, que restringe aún más las actividades para quienes habitan la zona.

Respecto a los actores identificados, se registraron en orden de importancia los siguientes: habitantes de la zona, el INDERENA (representado actualmente por Corpoboyacá, quien ejerce administración sobre la Reserva), la Policía Nacional, la Alcaldía Mayor de Tunja y, por último, la Junta de Acción Comunal y la Asociación de Productores de Leche.

Para el presente estudio se modificaron las 6 posibles dimensiones a identificar en los conflictos ambientales propuestas por Rodríguez (2016), esto con el fin de proporcionar un contexto más cercano al caso (figura 4). Así, se identifican las siguientes dimensiones:

- Temporal: conflicto existente desde 1976, agudizándose en la última década con el incremento de habitantes y actividades ganaderas en la zona.
- Biótica: el conflicto se desarrolla en un paisaje rural, donde la matriz está compuesta por un relicto de bosque altoandino de aproximadamente 159 ha, rodeado de parches de pasturas, cultivos de papa, zanahoria y legumbres, con corredores de cercas vivas compuestas por plantaciones de árboles maderables comerciales.

- Social: comunidad campesina de vocación agrícola y ganadera, la cual habita en el sector desde hace al menos tres generaciones. La formación académica de la población es escasa y la cabeza de hogar predominante es de género masculino.
- Económica: los habitantes de la zona obtienen sus ingresos principalmente de la actividad agropecuaria, a partir de la comercialización de productos y subproductos como la cosecha de cultivos, leche y queso y el pago de jornal por pastoreo de ganado. En su mayoría, los habitantes se encuentran bajo el régimen de salud subsidiado, lo que permite inferir que sus ingresos netos son mínimos. Por otra parte, los costos de administración y mantenimiento de la Reserva no son claros y demuestran baja efectividad de la inversión del tesoro público.
- Legal: el marco normativo del conflicto está dado por la Ley General Ambiental (Ley 99 de 1993), el Código Nacional de Recursos Naturales Renovables (Decreto-Ley 2811 de 1974), la Ley 1333 de 2009 (sancionatoria ambiental), el Código Nacional de Policía y Convivencia (Ley 1801 de 2016), el Decreto 2372 de 2010 (compilado en el Decreto único Reglamentario 1076 de 2015) y la Resolución 362 de 1976, Acuerdo 036 de 1976.

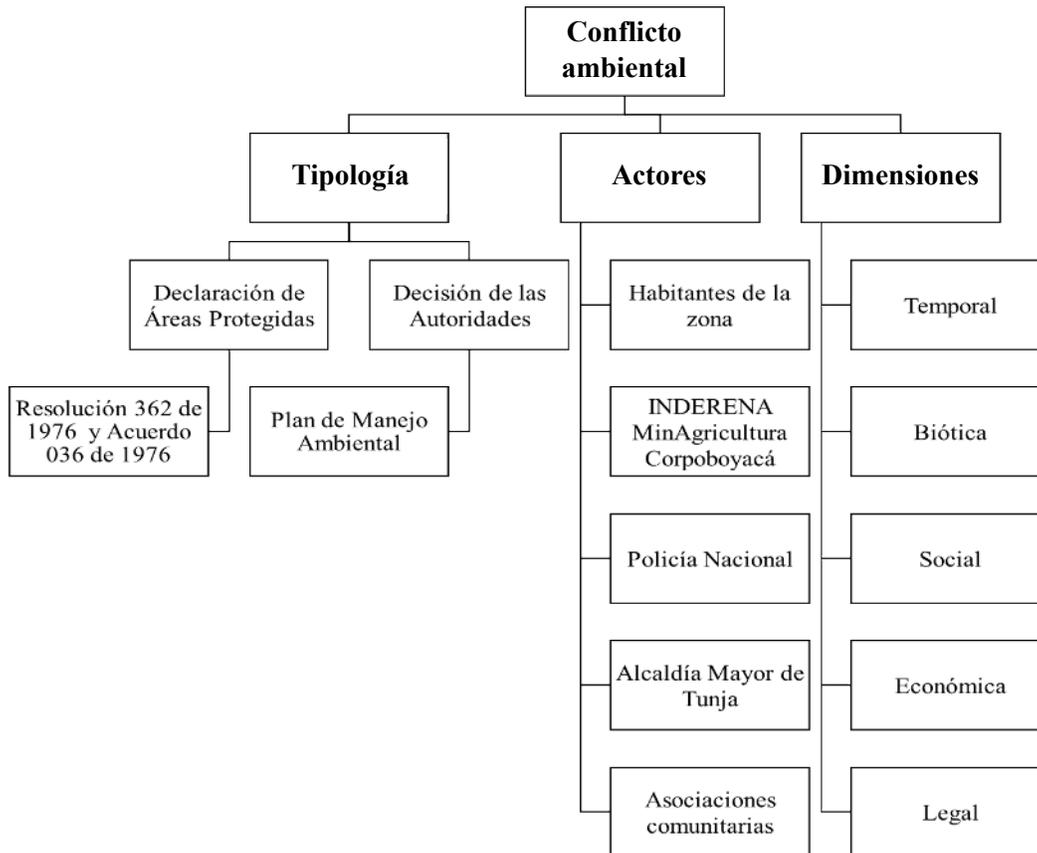
DISCUSIÓN

La primera cuestión por resolver es la vocación de los hogares entrevistados, pues esto permite entender la posición socioeconómica de los habitantes de la zona de influencia y su dependencia hacia la actividad ganadera. Siete de los ocho hogares entrevistados obtienen sus ingresos directamente de la venta diaria de leche y de la venta final del semoviente, así como del pago de jornales diarios en fincas aledañas con cultivos. Por lo tanto, se afirma que la población del área de influencia es netamente campesina, siguiendo la definición plasmada en la Declaración de las Naciones Unidas sobre los Derechos de los Campesinos (ONU, 2018).

Según los entrevistados, el asentamiento de la comunidad en el territorio surge mucho tiempo antes de la declaratoria. De igual modo, a pesar de reconocer que la creación de la Reserva se ejecutó hace 43 años,

afirman que durante este período no ha existido una autoridad ambiental competente que les informe los límites de la Reserva, ni las prohibiciones o permisos para ejercer actividades económicas dentro de la misma. Todo esto apunta a que el actor generador del conflicto es la autoridad ambiental, tanto la extinta (INDERENA) como la actual (Corpoboyacá).

Figura 4. Determinación del tipo de conflicto, los actores y las dimensiones del conflicto



Fuente: elaboración propia con base en la metodología de Rodríguez (2016).

Las implicaciones ecológicas del conflicto radican en el deterioro que generan los sistemas de ganadería en los sistemas naturales poco intervenidos. En primer lugar se encuentra la erosión hídrica y de masa, pues el tránsito de animales afecta la estabilidad del suelo y el flujo de agua, causando una pérdida acelerada e irreversible del suelo que conlleva a una disminución en la productividad de la vegetación a través del tiempo (Murgueitio, 2003). En segundo lugar, la deforestación es una de las consecuencias más visibles derivadas de la actividad ganadera, al punto de convertirse en uno de los factores más influyentes en el cambio del uso de suelo por cuenta de la ampliación de su frontera (Armenteras & Rodríguez, 2014). Así, se evidencia que esta actividad genera grandes afectaciones a los ecosistemas naturales, tales como su reducción, simplificación y vulnerabilidad por la pérdida irreversible de biodiversidad y las invasiones biológicas (FAO, 2019; Zuluaga *et al.*, 2011).

De acuerdo con los tipos de conflicto identificados en el área, esto es “declaración de áreas protegidas” y “decisiones de las autoridades”, es importante recordar que la figura de Reserva debe ser entendida como una restricción al uso del suelo para su utilización con fines netamente de conservación y uso conforme a las actividades definidas en la zonificación, mas no como una restricción al núcleo esencial del derecho de propiedad. Para el presente caso, la Reserva El Malmo está dividida en dominio mixto, es decir, 59,7 ha pertenecen al Estado y 100 ha a privados (Alcaldía Mayor de Tunja, 2016), por lo que el desconocimiento de los límites de la Reserva le impiden al campesino saber hasta dónde puede ejercer su derecho de uso y dónde empiezan las limitaciones al mismo.

De otra parte, los años de ausencia del PMA (pese a la insistencia de Corpoboyacá de actualizar un contenido que legalmente no existe), sumado a la ínfima mención del área protegida en el Plan de Ordenamiento Territorial expedido por la Alcaldía de Tunja, agregan un inconveniente mayor a la disputa: la ausencia de zonificación interna de la Reserva (zona de preservación, restauración, uso sostenible y uso público y sus correspondientes subzonas), de acuerdo con lo dispuesto en el artículo 34 del Decreto 2372 de 2010 (artículo 2.2.2.1.4.2. del Decreto Único Reglamentario 1076 de 2015), a lo que habría que añadir la omisión del establecimiento de función amortiguadora como determinante ambiental del ordenamiento territorial (2.2.2.1.3.10. del Decreto Único Reglamentario 1076 de 2015). Todo esto impide entender dónde deberían las autoridades aplicar medidas preventivas y correctivas.

Por lo anterior, las actuaciones de control por parte de Corpoboyacá, en su eje sancionatorio, y la Policía Nacional, en su división ambiental, pueden empeorar la situación al imponer medidas sancionatorias como la amonestación escrita, el decomiso de semovientes o multas, entre otras (artículo 36 y subsiguientes de la Ley 1333 de 2009 y artículo 103 de la Ley 1801 de 2016), sin la debida socialización de las razones que soportan las medidas, así como el contenido de las mismas.

El punto neurálgico del conflicto radica en la asimetría de poder entre los involucrados, configurada en la eliminación de la participación ciudadana en las decisiones adoptadas, toda vez que no tener voz ni voto en la forma en que se delimita el territorio aumenta la zozobra y predispone negativamente a los campesinos, quienes a pesar de tener voluntad de conservar mientras trabajan no conocen la forma correcta de hacerlo.

Una solución viable a este conflicto consiste en la elaboración y adopción de un PMA realmente concertado por las partes, donde todos los actores puedan aportar sobre sus necesidades y que estas sean completamente incorporadas en el documento final. Allí se deberían explicar de forma clara los límites de la Reserva, indicando qué actividades de uso de suelo están permitidas, así como las sanciones que puede acarrear el no cumplimiento de los compromisos adquiridos por los implicados.

Por otro lado, dicha concertación debe ser pública y de amplia difusión, con el fin de que todos los actores del conflicto estén debidamente informados y participen activamente. En este punto, las diferentes asociaciones comunitarias, como la Junta de Acción Comunal de la vereda Barón Germania, cumplen un papel importante al ejercer su función veedora, garantizando que se escuche la posición de todos los habitantes del sector y exigiendo acompañamiento de la autoridad ambiental y policial cuando se incumpla lo pactado. De igual modo, las asociaciones de productores lecheros también deberían solicitar constante capacitación sobre los impactos de la ganadería en la Reserva. Un excelente complemento a la solución aquí propuesta es el establecimiento de programas o esquemas de pago por servicios ambientales (PSA), los cuales han demostrado ser un instrumento idóneo para avanzar en materia socioeconómica y de derechos fundamentales para los habitantes de las áreas protegidas en Colombia (Guzmán & Ubajoa, 2018). El tipo de PSA que se podría establecer es el pago de un paquete de servicios que incluya la biodiversidad y la captura de carbono en sistemas silvopastoriles, esquema que se ha implementado en áreas de pasturas degradadas en países como Nicaragua, Costa Rica y Colombia, el cual se basa en determinar los usos del suelo y asociar cada uno de estos con un sistema de puntos que busca identificar los pagos respectivos y calcular el índice de servicio ambiental correspondiente (Pagiola *et al.*, 2004; Wunder *et al.*, 2007).

CONCLUSIONES

Las decisiones de las autoridades sobre los territorios rurales son generadores de conflictos ambientales cuando no se realizan concertaciones previas con la comunidad directamente afectada. La oportuna discusión y socialización del contenido de estas decisiones puede evitar la creación de nuevas disputas.

La ausencia del documento guía del área protegida (Plan de Manejo Ambiental) genera un limbo jurídico con respecto a las actividades permitidas y prohibidas en su interior, dificultando la aplicación de medidas correctivas y policivas por parte de las autoridades competentes. Por otra parte, y para el caso analizado, la Alcaldía Mayor de Tunja no tiene en cuenta el carácter de la Reserva como determinante ambiental en la actualización del Plan de Ordenamiento Territorial.

Las comunidades rurales suelen presentar predisposición negativa ante el actuar de las autoridades, al asimilarlas con la imagen de “cobradores y castigadores” en lugar de aliados en la administración de los recursos naturales. Este es el caso de la población del área de influencia de la Reserva Forestal Protectora El Malmo, la cual no reconoce positivamente a la Corporación Autónoma Regional de Boyacá (Corpoboyacá) ni a la Alcaldía Mayor de Tunja como instituciones promotoras de acciones de conservación, sino como entes meramente vigilantes.

Es urgente que las autoridades ambientales y locales realicen la zonificación que el PMA exige de una manera participativa y concertada, donde los intereses de todos los actores se plasmen en iguales proporciones. Así mismo, se debe educar a la comunidad sobre las restricciones y la responsabilidad que la adopción de dicho documento conlleva, ofreciendo también alternativas de producción sostenible que garanticen su derecho al mínimo vital.

REFERENCIAS

- Alcaldía Mayor de Tunja (2016). *Plan de desarrollo municipal 2016-2019*. https://alcaldiatunja.micolombiadigital.gov.co/sites/alcaldiatunja/content/files/000354/17668_conceptoctppdm20162019version0329.pdf
- Álvarez, G. L. (2011). *Las áreas protegidas en Colombia*. Universidad Externado de Colombia.
- Andean Geological Services [AGS]. (2018). *Plan de Manejo Reserva Forestal del Malmo*. Corpoboyacá.
- Armenteras, D., & Rodríguez, N. (2014). Dinámicas y causas de deforestación en bosques de Latinoamérica: una revisión desde 1990. *Colomb Forest.*, 17(2), 233-246. <https://doi.org/10.14483/udistrital.jour.colomb.for.2014.2.a07>
- Congreso de la República de Colombia (1993, 22 de diciembre). *Ley 99 de 1993. Por la cual se crea el Ministerio del Medio Ambiente, se reordena el sector público encargado de la gestión y conservación del medio ambiente y los recursos naturales renovables, se reorganiza el Sistema Nacional Ambiental, SINA y se dictan otras disposiciones*. Diario Oficial N.º 41146.
- Congreso de la República de Colombia (2009, 21 de julio). *Ley 1333 de 2009. Por la cual se establece el procedimiento sancionatorio ambiental y se dictan otras disposiciones*. Diario Oficial N.º 47417.
- Congreso de la República de Colombia (2016, 29 de julio). *Ley 1801 de 2016. Por la cual se expide el Código Nacional de Seguridad y Convivencia Ciudadana*. Diario Oficial N.º 49949.
- Cuenca-Jiménez, N. J., Chavarro-Miranda, F., & Díaz-Gantiva, O. H. (2008). El sector de la ganadería bovina en Colombia. Aplicación de modelos de series de tiempo al inventario ganadero. *Rev. Fac. Cienc. Econ.*, 16(1), 165-177.
- Departamento Administrativo Nacional de Estadística [DANE]. (2016). *3er Censo Nacional Agropecuario: Hay campo para todos. Tomo 2 - Resultados*. DANE. <https://www.dane.gov.co/files/images/foros/foro-de-entrega-de-resultados-y-cierre-3-censo-nacional-agropecuario/CNATomo2-Resultados.pdf>
- Doran, J., & Parkin, T. (1994). Defining and assessing soil quality. En J. W. Doran, D. C. Coleman, D. F. Bezdicek, & B. A. Stewart (eds.), *Defining soil quality for a sustainable environment* (pp. 3-21). American Society of Agronomy. <https://doi.org/10.2136/sssaspepub35>
- Environmental System Research Institute [ESRI]. (2015). *ArcGIS desktop* (Versión 10.3.1) [Software]. <https://www.esri.com/es-es/arcgis/about-arcgis/overview>
- Guzmán, L., & Ubajoa, J. (2018). *El pago por servicios ambientales como un instrumento idóneo para gestionar las aguas y otros recursos naturales, así como para avanzar en materia socioeconómica*. Universidad Externado de Colombia.
- Instituto Nacional de los Recursos Naturales Renovables y del Ambiente [INDERENA] (1976, octubre 23). *Acuerdo 036 de 1976. Por el cual se declara Área de Reserva Forestal Protectora la zona denominada EL MALMO, ubicada en la vereda Barón Germania, en el Municipio de Tunja*. INDERENA.

- Lal, S., Safriel, U., & Boer, B. (2012). *Zero net land degradation: A new sustainable development goal for Rio+ 20*. http://catalogue.unccd.int/991_Zero_Net_Land_Degradation_Report_UNCCD_May_2012.pdf
- Ministerio de Agricultura de Colombia (1976, 17 de diciembre). *Resolución 362 de 1976 Por la cual se aprueba un acuerdo de la Junta Directiva del Instituto Nacional de los Recursos Naturales Renovables y del Ambiente-INDERENA*. MinAgricultura.
- Ministerio de Ambiente de Colombia (2012). *Política Nacional para la Gestión Integral de la Biodiversidad y sus Servicios Ecosistémicos (PNGIBSE)*. Instituto Humboldt.
- Ministerio de Ambiente y Desarrollo Sostenible de Colombia (2016, 28 de octubre). *Resolución 1768 de 2016. Por la cual se delimita el páramo Rabanal - Río Bogotá y se adoptan otras determinaciones*. MinAmbiente.
- Mora-Marín, M. A., Ríos-Pescador, L., Ríos-Ramos, L., & Almario-Charry, J. L. (2017). Impacto de la actividad ganadera sobre el suelo en Colombia. *Ingeniería y Región*, 17, 1-12. <https://doi.org/10.25054/22161325.1212>
- Murgueitio, E. (1999). Reconversión ambiental y social de la ganadería en Colombia. *Revista Mundial de Zootecnia*, 93, 2-15.
- Murgueitio, E. (2003). Impacto ambiental de la ganadería de leche en Colombia y alternativas de solución. *Livest. Res. Rural Dev.*, 15(10). <https://lrrd.cipav.org.co/lrrd15/10/murg1510.htm>
- Murgueitio, E., & Ibrahim, M. (2008). Ganadería y medio ambiente en América Latina. En E. Murgueitio, C. Cuartas, & J. Naranjo (Ed.), *Ganadería del Futuro: Investigación para el Desarrollo* (pp. 19-39). Fundación CIPAV.
- Organización de las Naciones Unidas [ONU]. (2018). *Declaración de las Naciones Unidas sobre los derechos de los campesinos y de otras personas que trabajan en las zonas rurales*. https://digitallibrary.un.org/record/1650694/files/A_HRC_RES_39_12-ES.pdf
- Organización de las Naciones Unidas para la Alimentación y la Agricultura [FAO]. (2019, marzo). *Ganadería sostenible y cambio climático en América Latina y el Caribe*. <http://www.fao.org/americas/prioridades/ganaderia-sostenible/es/>
- Pagiola, S., Agostini, P., Gobbi, J., de-Haan, C., Ibrahim, M., Murgueitio, E., Ramírez, E., Rosales, M., & Ruíz, J. P. (2004). *Pago por servicios de conservación de la biodiversidad en paisajes agropecuarios* (Environment Department Papers). The World Bank.
- Presidencia de la República de Colombia (1975, 27 de enero). *Decreto 2811 de 1974. Por el cual se dicta el Código Nacional de Recursos Naturales Renovables y de Protección al Medio Ambiente*. Diario Oficial N.º 34243.
- Presidencia de la República de Colombia (2010, 1 de julio). *Decreto 2372 de 2010 Por el cual se reglamenta el Decreto-ley 2811 de 1974, la Ley 99 de 1993, la Ley 165 de 1994 y el Decreto-ley 216 de 2003, en relación con el Sistema Nacional de Áreas Protegidas, las categorías de manejo que lo conforman y se dictan otras disposiciones*. Diario Oficial N.º 47757.
- Presidencia de la República de Colombia (2013, 12 de septiembre). *Decreto 1985 de 2013. Por el cual se modifica la estructura del Ministerio de Agricultura y Desarrollo Rural y se determinan las funciones de sus dependencias*. Diario Oficial N.º 48911.
- Presidencia de la República de Colombia (2015, 26 de mayo). *Decreto 1076 de 2015. Por medio del cual se expide el Decreto Único reglamentario del sector ambiente y desarrollo sostenible*. Diario Oficial N.º 49523.
- Ramos, A., & Armenteras, D. (2019). Interceptación y escorrentía del bosque altoandino en la reserva forestal protectora "El Malmo". *Acta Biol. Colomb.*, 24(1), 97-108. <https://doi.org/10.15446/abc.v23n3.67039>
- Rodríguez, G. (2016). *Los conflictos ambientales en Colombia y su incidencia en los territorios indígenas*. Universidad del Rosario. <https://doi.org/10.12804/tj9789587387407>

- Vergara-Vergara, W. (2010). La ganadería extensiva y el problema agrario. El reto de un modelo de desarrollo rural sustentable para Colombia. *Revista Ciencia Animal*, 3, 45-53.
- Wunder, S., Wertz-Kanounnikoff, S., & Sánchez, R. M. (2007). Pago por servicios ambientales: una nueva forma de conservar la biodiversidad. *Gaceta ecológica*, 84, 39-52.
- Zuluaga, A., Giraldo, C., & Chará, J. (2011). *Servicios ambientales que proveen los sistemas silvopastoriles y los beneficios para la biodiversidad. Manual 4, Proyecto Ganadería Colombiana Sostenible*. <http://www.cipav.org.co/pdf/4.Servicios.Ambientales.pdf>

Detección de contratistas multiobjeto mediante minería de textos para focalizar el ejercicio del control y vigilancia fiscal

Detection of Multi-object Contractors through Text Mining to Targeting the Exercise of Fiscal Control and Surveillance

Manuel Francisco Dulce Vanegas^{ad}, Adam Beltrán Gómez^{bce}

^a Maestría en Ingeniería y Analítica de datos, Facultad de Ciencias Naturales e Ingeniería, Universidad de Bogotá Jorge Tadeo Lozano, Colombia

^b Facultad de Ingeniería, Universidad de Cundinamarca, Colombia

^c Facultad de Ingeniería, Universidad Católica de Colombia, Colombia

^d manuel.dulcev@utadeo.edu.co | <https://orcid.org/0000-0002-8210-7754>

^e <https://orcid.org/0000-0001-8329-6991>

RESUMEN

Las entidades fiscalizadoras superiores, y en específico su ente rector, la Organización Internacional de las Entidades Fiscalizadoras Superiores (INTOSAI), han impulsado en los últimos cuatro años iniciativas encaminadas al uso de tecnologías y métodos para sus procesos de vigilancia y fiscalización que sean replicables y que generen resultados tangibles en el contexto fiscal. En este sentido, la Contraloría General de la República de Colombia viene fortaleciendo su infraestructura tecnológica y capacidades técnicas con miras a mejorar y optimizar sus esfuerzos en cuanto a la vigilancia de los recursos de los colombianos. Aunque dicha tarea no es sencilla, esta entidad ha logrado detectar patrones de aquellos contratistas que acaparan la contratación estatal, logrando estar en diferentes sectores económicos sin tener probablemente la competencia técnica para cumplir el objeto contractual estipulado. A estos se les conoce en el ámbito de la Contraloría General como contratistas “multiobjeto”. En el presente artículo

se muestra la construcción de un conjunto de datos de 1.998 registros etiquetado por expertos, que corresponden a contratos del sector educativo en Colombia. Con este instrumento se llevó a cabo el entrenamiento y las pruebas sobre un clasificador automático construido para los objetos contractuales a fin de detectar presuntos contratistas “multiobjeto”. Adicionalmente, se encontró que el mejor algoritmo de clasificación fue “Máquina de Soporte Vectorial Lineal”, con una exactitud de 84 %, el cual permitió finalmente listar por agrupamiento los presuntos contratistas de este tipo.

Citation: Dulce-Vanegas, M. F. y Beltrán-Gómez, A. (2021). Detección de contratistas multiobjeto mediante minería de textos para focalizar el ejercicio del control y vigilancia fiscal. *Mutis*, 11(1), 48-64. <https://doi.org/10.21789/22561498.1732>

Recibido: Noviembre 18, 2020. **Aceptado:** Marzo 15, 2021.

Copyright: ©2021. Chávez-Plazas, Y. A., Ramírez-Mahecha, M. C. y Camacho-Kurmen, J. E. (2021). This is an open-access article, which permits unrestricted use, distributions and reproduction in any medium, provided the original author and source are credited.

Competing Interests: The authors have no conflict of interest.

Palabras clave: minería de texto, aprendizaje de máquina, contratación estatal, control fiscal.



Abstract

Supreme audit institutions, and specifically its governing body, the International Organization of Supreme Audit Institutions (INTOSAI), have promoted during the last four years a series of initiatives in the fiscal context aimed at the use of technologies and methods that are replicable and generate tangible results, thus reinforcing the surveillance and auditing processes carried out by supreme audit institutions. In this sense, the Comptroller General of the Republic of Colombia has been strengthening its technological infrastructure and technical capacities in order to improve and optimize its efforts in the monitoring of the resources of Colombian citizens. Although this task is not an easy one, this entity has managed to detect patterns of contractors who monopolize state contracting and are inserted into different economic sectors, without probably having the technical competence to fulfill stipulated contractual deeds. These subjects are known in the field of the General Comptroller's office as "multi-object" contractors. This article explains the construction of a data set of 1,998 records labeled by experts that correspond to education sector contracts. Training and tests were carried out with this tool on an automatic classifier built for the contractual objects in order to detect suspected "multi-object" contractors. It was found that the best classification algorithm was the "Linear Vector Support Machine," with an accuracy of 84%, which will eventually find presumed multi-object contractors by grouping.

Keywords: Text mining, machine learning, state procurement, fiscal control and surveillance.

INTRODUCCIÓN

Entre los diferentes procedimientos para la detección del fraude, en el marco de la Séptima Conferencia de Criminología Financiera (Othman *et al.*, 2015) se manifestó que las operaciones de auditoría siguen siendo las técnicas más relevantes para su detección, lo cual revela la importancia del rol de los auditores en la detección de este fenómeno. Aunque las operaciones de auditoría se encuentren supeditadas a la legislación o las características sociodemográficas de cada país, existen métodos aplicables a diferentes contextos donde el proceso auditor requiera realizar la búsqueda de patrones relacionados con la incidencia de actividades fraudulentas. Entre estos se encuentra la minería de datos, que permite realizar clasificaciones, clusterización y segmentación de datos de manera semiautomatizada, como ocurrió en Rumania para detectar patrones de fraude en el sistema de salud de este país (Bologa *et al.*, 2010).

Mientras tanto, la Organización Internacional de las Entidades Fiscalizadoras Superiores (INTOSAI), ente rector de las entidades fiscalizadoras superiores (EFS), cuyo objetivo principal es "promover el intercambio de ideas, experiencias y conocimientos entre las EFS de países alrededor del mundo" (INTOSAI, s. f.), realiza importantes esfuerzos para facilitar el acceso y aprovechamiento de las iniciativas de desarrollo de capacidades para que las EFS cumplan con su labor en cada

nación. Según Harib Saeed Al-Amimi, presidente de INTOSAI, "la automatización robótica de procesos tiene el potencial de consumir trabajo de auditoría repetitivo y hacerlo de manera más precisa, confiable e incansable, en una fracción del tiempo" (2020, p. 5).

Debido a lo anterior, se han suscitado iniciativas en esta materia dentro de las EFS, tales como la identificación de riesgos en la ejecución de los recursos públicos asociados bien sea al contrato, la entidad pública contratante o al contratista. Por ejemplo, la Contraloría de Bogotá (2018), entidad fiscalizadora territorial, expuso un ejercicio de uso de técnicas de minería de datos en las labores de auditoría fiscal durante el Concurso Regional sobre Buena Gobernanza, edición 2018 (Giraldo-Polanía *et al.*, 2018). En dicho ejercicio, la Contraloría de Bogotá logró establecer un proceso de responsabilidad fiscal por USD 72.000 en dos hallazgos, tras identificar fraude en la asignación de bonos de alimentos para las vigencias 2016 y 2017 mediante el cruce de información y la extracción de textos.

De igual forma, la Contraloría General de la República de Colombia (CGR), ente superior de control fiscal a nivel nacional, dio a conocer en 2019 la iniciativa Océano (ahora Dirección de Análisis, Información y Reacción Inmediata), que surge con el fin de brindar una solución a la brecha entre la labor de vigilancia

en la correcta ejecución de los recursos públicos y la capacidad tecnológica para poder cumplir con su labor constitucional. Dentro de los logros obtenidos, se pudo establecer con certeza las cifras de contratación, los riesgos del proceso (mediante mallas empresariales) y los cruces con otras fuentes de información, con lo cual ha sido posible establecer inhabilidades de contratación y analizar a los contratistas multiobjeto, que como indica Carlos Felipe Córdoba Larrarte (2019), contralor general de la república, son aquellos contratistas que reflejan patrones en la forma de asociarse temporalmente con personas o empresas con experiencia para conseguir contratos en diferentes sectores, por lo cual generan riesgos de posible cartelización, atrasos en la ejecución y, en últimas, redundan en incumplimientos contractuales.

A pesar de los esfuerzos en cuanto a la detección de fraudes antes mencionados, aún no existen trabajos que automaticen la identificación de contratistas multiobjeto. Por este motivo, el presente trabajo aporta al proceso de detección de este tipo de contratistas, con el fin de facilitar la labor del proceso auditor mediante un modelo multiclase. Lo anterior se encuentra soportado en estudios que sugieren que “la detección del fraude mediante el empleo de un proceso automático permite clasificar de forma masiva operaciones o sujetos e identificar casos de alto riesgo” (Álvarez-Jareño *et al.*, 2018, p. 4), basados principalmente en cuatro pasos: la detección del fraude, la investigación del fraude, la confirmación de este y su prevención, lo cual ha sido fundamental para frenar procesos de fraude y corrupción en entidades financieras o gubernamentales, incluso antes de que ocurran.

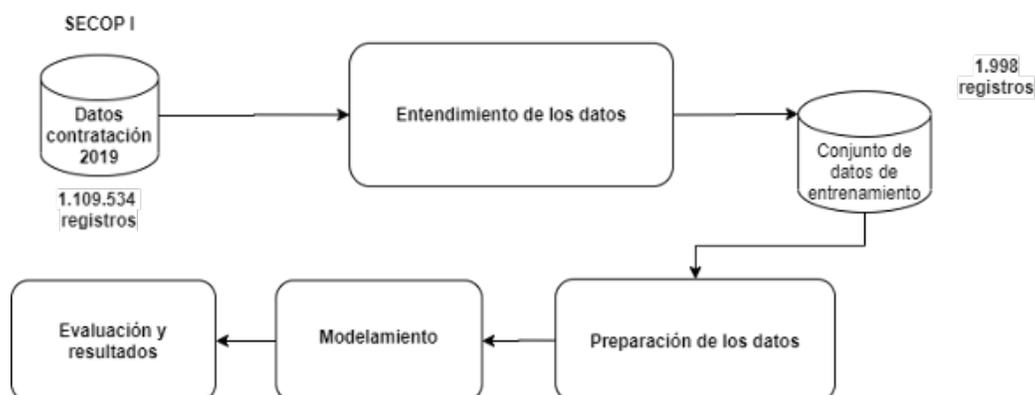
Para lograr el propósito mencionado, en esta investigación se utilizaron los algoritmos de mayor uso en la clasificación multiclase, que según Mohamed (2005) son: árboles de decisión, k -vecinos cercanos, Bayes ingenuo, máquinas de soporte vectorial y redes neuronales perceptrón multicapa. Los árboles de decisión (Song & Lu, 2015) permiten clasificar los ejemplos, tomándolos principalmente desde la raíz hasta algún nodo hoja y particionando los datos en función del valor del atributo. K -vecinos cercanos (García *et al.*, 2018) es un algoritmo de clasificación que usa los datos directamente para la clasificación, por lo que no se hace necesario considerar detalles de la construcción del modelo para estimar la pertenencia o no en una clase

puntual, siendo la cantidad de vecinos cercanos k su único parámetro ajustable. Por su parte, Bayes ingenuo multinomial (García *et al.*, 2018) es un clasificador probabilístico basado en la predicción de la probabilidad de posibles resultados en una categoría, para los que se requieren ejemplos previamente definidos. De otro lado, máquinas de soporte vectorial (Rennie & Rifkin) son algoritmos de aprendizaje supervisado para problemas de regresión y reconocimiento de patrones que construyen su solución basándose en un subconjunto de los datos de entrenamiento; estos se clasifican en SVM lineal y no lineal, donde el modelo lineal parte de un conjunto de puntos etiquetados para entrenamiento linealmente separable y el no lineal resuelve el problema de los datos no separables hasta cierto grado, guardando un margen de error en la clasificación. Por último, las redes neuronales de tipo perceptrón multicapa (Hsu, 2020) son un tipo especial de red neuronal en el que se apilan varias capas de perceptrones, que generan una predicción para una entrada de acuerdo con una función de activación.

En cuanto a la estructura del presente documento, la sección después de esta introducción muestra el uso del marco metodológico CRISP-DM bajo el cual se realizó el entendimiento de los datos, así como el preprocesamiento y la construcción del modelo. Luego, se muestran los resultados del ejercicio realizado, donde se compara el desempeño de los diferentes algoritmos utilizados. En la sección posterior se presenta la discusión respecto a las dificultades en los datos, las categorías y la interpretación que el modelo tiene respecto a los mismos. Finalmente, se formulan las conclusiones del ejercicio, las lecciones aprendidas y los aspectos a tener en cuenta para el mantenimiento y la mejora del presente modelo.

MATERIALES Y MÉTODOS

La metodología utilizada para el desarrollo del presente proyecto fue CRISP-DM. Debido al uso generalizado de esta metodología en gran parte de los proyectos de minería de datos, “el modelo de proceso CRISP-DM tiene como objetivo hacer grandes proyectos de minería de datos, menos costosos, más confiables, repetibles, manejables y de rápida ejecución” (Wirth & Hipp, 2000, p. 2). En cada fase explicaremos las actividades desarrolladas para este proyecto, las cuales se encuentran resumidas en la figura 1.

Figura 1. Metodología del proyecto

Fuente: elaboración propia.

Gran parte del desarrollo se fundamentó en el ejercicio de “clasificación de las quejas de financiación del consumidor” (Li, 2018), el cual permite realizar una clasificación multiclase a partir del texto de las quejas de clientes y la categorización del producto financiero referido. Este ejercicio puede ser replicable, salvo en algunos aspectos del negocio propios del contexto del control fiscal.

En cuanto a las herramientas, se utilizó Python 3 en Google Colab, con el uso de bibliotecas como scikit-learn (Pedregosa *et al.*, 2011), la cual tiene un ecosistema de algoritmos de clasificación propicio para el desarrollo del proyecto, y NLTK para el entendimiento de lenguaje natural en el procesamiento y extracción de los textos de los objetos contractuales. En la fase de exploración de los datos también se utilizó la herramienta de minería de datos llamada KNIME, que permite la integración y realización de pruebas en la construcción del conjunto de datos preentrenado.

Entendimiento del negocio

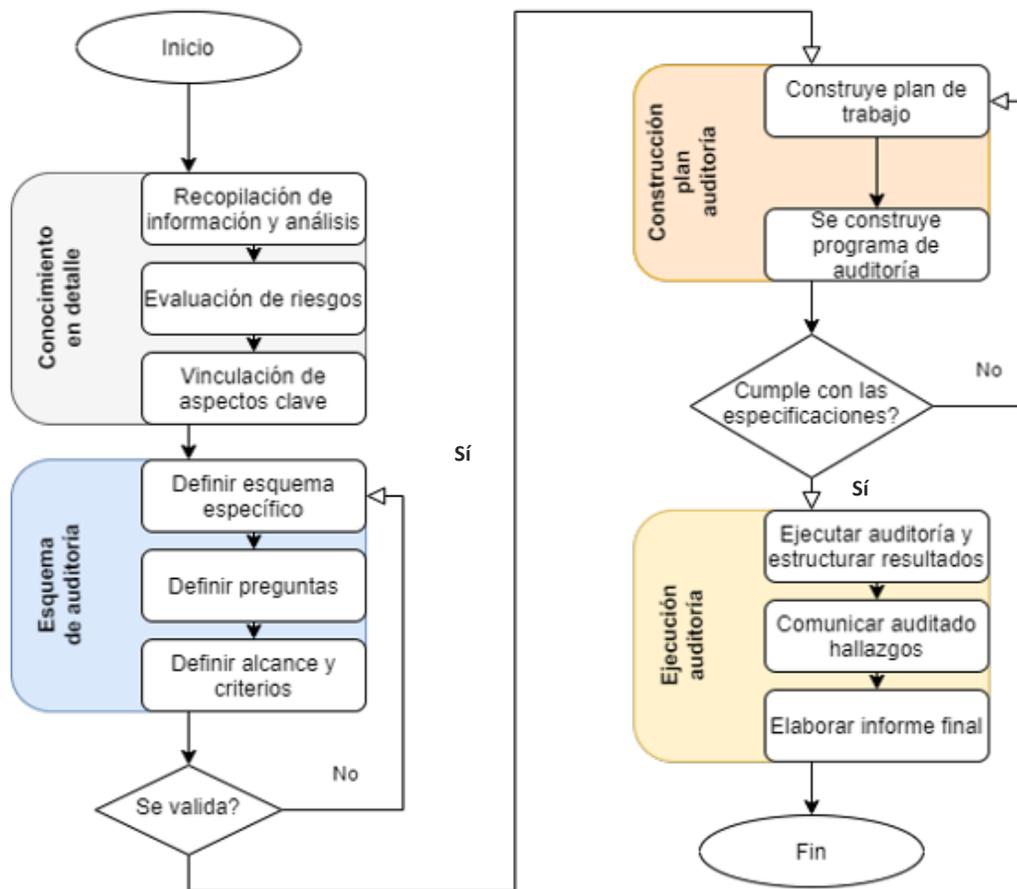
En esta primera fase se brinda un contexto más amplio del estado en el que se encuentra la CGR en materia de implementación de tecnologías para el seguimiento de la contratación, así como de la manera en que este proyecto responde a la necesidad de la entidad dentro de su proceso auditor (control fiscal micro).

La CGR (2018) plantea en el objetivo 5 de su plan estratégico la búsqueda del fortalecimiento de las capacidades y los servicios tecnológicos con miras a

impulsar la transformación digital. A partir de ello, la entidad ha implementado el uso de tecnologías disruptivas con el fin de analizar información masivamente y detectar patrones de comportamiento atípico que generen riesgos en la contratación y ejecución presupuestal del recurso público. En la información recopilada para sus procesos de auditoría fiscal, la CGR ha logrado recopilar varias fuentes de información contractual para contar con una base de datos más cercana a la realidad, para lo cual es necesario involucrar técnicas avanzadas de minería de datos que permitan generar valor y brindar agilidad a los procesos auditores.

En la figura 2 se observa de manera resumida el proceso auditor que realiza la CGR como ente de fiscalización, el cual consiste en la verificación de la gestión financiera, contable, administrativa y contractual de las entidades sujetas a vigilancia por parte de este órgano de control. Como en todo proceso de auditoría, es necesario que el equipo auditor conozca y planee la auditoría teniendo previo conocimiento de la información con la que cuenta la entidad a inspeccionar, la evaluación de riesgos y el alcance de la auditoría; fase que culmina con la construcción del plan y el programa de auditoría. Como siguiente etapa, se realiza la auditoría *in situ* de acuerdo con la planeación establecida y siguiendo las guías de auditoría de la CGR, culminando con la identificación de hallazgos y la elaboración del informe final. En este contexto, el aporte de este proyecto se fundamenta en brindar las herramientas necesarias para la fase de planeación, específicamente en cuanto al conocimiento en detalle y el esquema de auditoría requerido para los contratistas multiobjeto.

Figura 2. Diagrama resumen del proceso auditor en la CGR



Fuente: elaboración propia.

Para realizar lo anterior es necesario clasificar los objetos contractuales de los registros de contratación del Estado colombiano reportados en la fuente SECOP I (Sistema Electrónico de Contratación Pública) para el sector de Educación, disponible en la página web de Colombia Compra Eficiente.¹ Paso seguido, mediante la puesta en práctica de técnicas de minería y clasificación de textos, se busca identificar los contratistas que tienen la característica de multiobjeto, es decir, aquellos que en la clasificación resultante tengan dos o más categorías asignadas.

Así, para el desarrollo de esta investigación se propusieron las siguientes actividades principales:

- Identificar las categorías óptimas de clasificación que se asignen a los objetos contractuales.

- Realizar el comparativo de validación de modelos de clasificación de texto para determinar el que más generalice la realidad.
- Ejecutar el modelo seleccionado y clasificar los objetos contractuales.
- Establecer los contratistas con multiplicidad de objetos.

Entendimiento de los datos

En esta segunda fase se describe el proceso de obtención y análisis preliminar de los datos que se utilizaron en el ejercicio. De igual forma, se realiza la identificación de las categorías utilizadas en el conjunto de datos de entrenamiento.

Se realizó la descarga de 1.109.534 registros de contratos suscritos en 2019 desde la plataforma de datos

1 <https://www.colombiacompra.gov.co/>

abiertos Colombia Compra Eficiente, en archivo plano (csv). Dicho archivo tiene un total de 71 atributos (tabla 1), entre los que se incluyen las variables de interés de nuestro ejercicio:

- *UID* (tipo texto): valor compuesto para identificar de manera individual cada registro.
- *Nombre de la Entidad* (tipo texto): nombre de la entidad del estado a la que corresponde el proceso, de la cual seleccionaremos los registros pertenecientes a entidades del sector educación.
- *Detalle del Objeto a Contratar* (tipo texto): detalle de la definición del bien o servicio a adquirir dentro del proceso.

Posteriormente, se seleccionaron los registros del sector Educación (10.120 registros), como se muestra en la figura 3.

Tabla 1. Atributos del conjunto de datos de entrada y atributos de interés

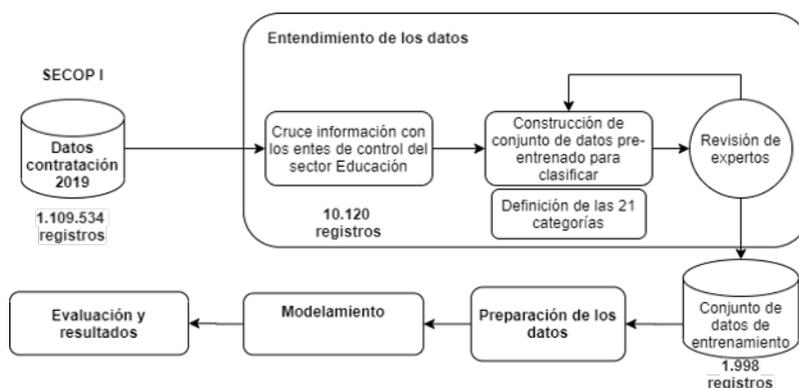
Nombre campo (71)	Tipo
UID	Cadena
Anno Cargue SECOP	Número (entero)
Anno Firma del Contrato	Número (entero)
Nivel Entidad	Cadena
Orden Entidad	Cadena
Nombre de la Entidad	Cadena
NIT de la Entidad	Cadena
Código de la Entidad	Número (entero)
ID Tipo de Proceso	Número (entero)
Tipo de Proceso	Cadena
Estado del Proceso	Cadena
Causal de Otras Formas de Contratación Directa	Cadena
ID de Regimen de Contratación	Número (entero)
Regimen de Contratación	Cadena
ID Objeto a Contratar	Número (entero)
Objeto a Contratar	Cadena
Detalle del Objeto a Contratar	Cadena
Tipo de Contrato	Cadena
Municipio Obtención	Cadena
Municipio Entrega	Cadena
Municipios Ejecución	Cadena
Fecha de Cargue en el SECOP	Cadena
Número de Constancia	Cadena

Nombre campo (71)	Tipo
Número de Proceso	Cadena
Número del Contrato	Cadena
Cuantía Proceso	Cadena
ID Grupo	Cadena
Nombre Grupo	Cadena
ID Familia	Número (entero)
Nombre Familia	Cadena
ID Clase	Número (entero)
Nombre Clase	Cadena
ID Ajudicación	Número (entero)
Tipo Identificación del Contratista	Cadena
Identificación del Contratista	Cadena
Nombre Razón Social Contratista	Cadena
Departamento y Municipio Contratista	Cadena
Tipo Documento Representante Legal	Cadena
Identificación del Representante Legal	Cadena
Nombre del Representante Legal	Cadena
Fecha de Firma del Contrato	Cadena
Fecha Ini Ejecución Contrato	Cadena
Plazo de Ejecución del Contrato	Número (entero)
Rango de Ejecución del Contrato	Cadena
Tiempo Adiciones en Dias	Número (entero)
Tiempo Adiciones en Meses	Número (entero)
Fecha Fin Ejecución Contrato	Cadena
Compromiso Presupuestal	Número (entero)
Cuantía Contrato	Cadena
Valor Total de Adiciones	Cadena
Valor Contrato con Adiciones	Cadena
Objeto del Contrato a la Firma	Cadena
ID Origen de los Recursos	Número (entero)
Origen de los Recursos	Cadena
Código BPIN	Número (entero)
Proponentes Seleccionados	Cadena
Calificación Definitiva	Cadena
ID Sub Unidad Ejecutora	Cadena
Nombre Sub Unidad Ejecutora	Cadena
Ruta Proceso en SECOP I	Cadena
Moneda	Cadena
EsPostConflicto	Cadena
Marcación Adiciones	Cadena
Posición Rubro	Cadena
Nombre Rubro	Cadena
Valor Rubro	Cadena

Nombre campo (71)	Tipo
Sexo Rep Legal Entidad	Cadena
Pilar Acuerdo Paz	Cadena
Punto Acuerdo Paz	Cadena
Municipio Entidad	Cadena
Departamento Entidad	Cadena

Fuente: elaboración propia.

Figura 3. Esquema de entendimiento de los datos



Fuente: elaboración propia.

Mediante juicio de expertos, se realizaron 3 iteraciones para discutir y establecer cuáles serían las categorías óptimas para clasificar los objetos en el sector educación. En la primera iteración se establecieron 6 categorías principales. En las siguientes iteraciones se determinó que el modelo tendría resultados más acertados al utilizar un segundo nivel, para así reducir la probabilidad de que el clasificador tuviese ambigüedades debido a la existencia de clases demasiado genéricas (por ejemplo, para el caso de los contratos de prestación de servicios, como se explica en este apartado). La tabla 2 muestra las 21 categorías definitivas para el desarrollo del modelo.

Tabla 2. Categorías iniciales y subcategorías definitivas

Categorías Iniciales	Subcategorías Derivadas
Arrendamiento	Arriendo inmueble
	Otros arriendos
Construcción y adecuaciones	Interventoría
	Obra civil
Convenios interadministrativos	Concurso público
	Convenio gratuidad educativa
	Convenios en actividades deportivas
	Otros convenios

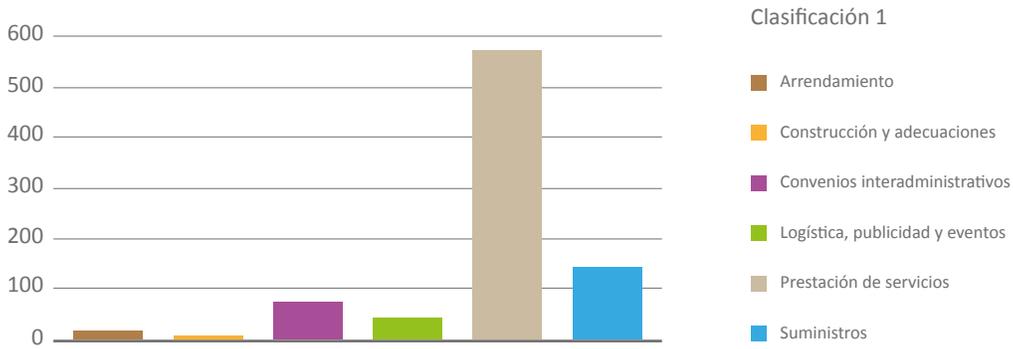
Categorías Iniciales	Subcategorías Derivadas
Logística, publicidad y eventos	Formación y capacitación
	Logística y transporte
Prestación de servicios	Apoyo jurídico
	Prestación de servicios
	Servicios auxiliares, técnicos y de apoyo a la gestión
	Servicios de alimentación
	Servicios de docencia
	Servicios de mantenimiento
	Servicios profesionales
Suministros	Elementos oficina y aseo
	Otros suministros
	Software y equipos de cómputo
	Compraventa

Fuente: elaboración propia.

Se seleccionó una muestra de 1.000 registros de los 10.120 disponibles para el sector analizado. Esta muestra fue validada en las primeras 6 categorías generales por medio de expresiones regulares y palabras clave en cada una de ellas. Sin embargo, durante el análisis y juicio de expertos se logró determinar una

coincidencia de 78 % en la clasificación humana respecto a las reglas de expresiones regulares, identificada por los 3 expertos a cargo. Además, se detectó que las ambigüedades entre los anotadores se encuentran sesgadas hacia prestación de servicios o suministros, por lo cual se hizo necesario subcategorizar esas dos clases a fin de reducir el sesgo y balancear la muestra (figura 4).

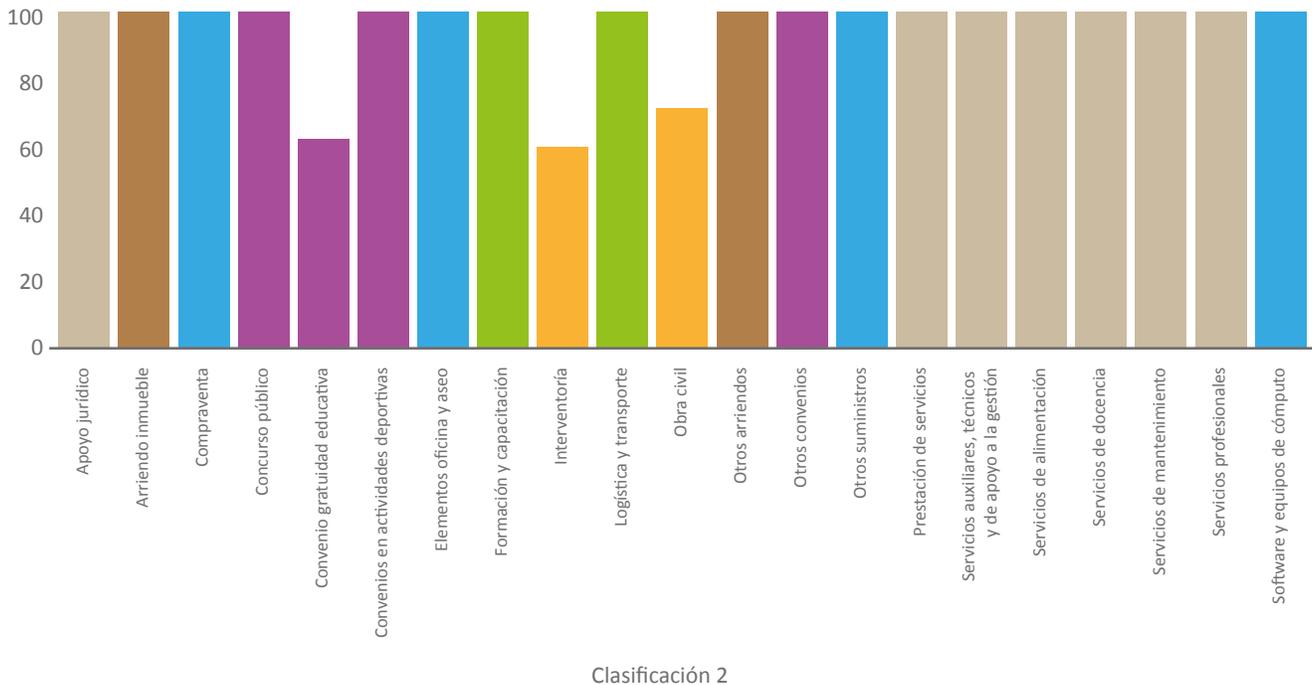
Figura 4. Distribución de los datos para las 6 categorías en primer nivel



Fuente: elaboración propia en Python.

Debido a lo anterior, se optó por realizar subcategorías junto con el juicio de expertos. A partir de las categorías principales se definieron subcategorías que se tomarán para reducir ambigüedades en la clasificación. Para esto se tomó una muestra de 100 registros por subcategoría, cuando la cantidad de registros así lo permitió, los cuales fueron clasificados manualmente e incluidos en el modelo, como se muestra en la figura 5.

Figura 5. Distribución de los datos para las 21 categorías en segundo nivel



Fuente: elaboración propia en Python.

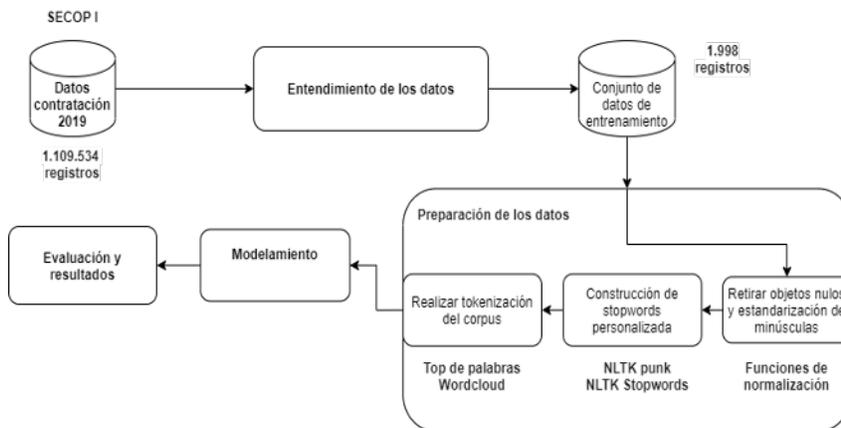
Finalmente, el conjunto de datos construido para el entrenamiento de los modelos se compone de 1.998 registros con 16 campos, de los cuales 3 serán seleccionados para la construcción del modelo: el objeto contractual (variable de entrada), clasificación2 (variable objetivo) y el uid (identificador único para integrar al conjunto de datos nuevamente y lograr identificar los contratistas multiobjeto).

Preparación de los datos

En esta fase se realiza la limpieza y mejora de la calidad de los datos incluidos en el corpus. El procesamiento de los datos, aspecto fundamental en el desarrollo de este proyecto, se muestra de manera

resumida en la figura 6. Este incluyó la construcción de funciones iterables a través de expresiones regulares y el uso de otras expresiones, contenidas en la librería NLTK y Scikit-learn, para la limpieza y el preprocesamiento de los datos en nuestra variable de entrada, con lo que se retiraron principalmente objetos contractuales vacíos, números, tildes, diéresis y espacios en blanco, además de reducir las letras minúsculas. La construcción de la lista de las palabras vacías en español se realizó tomando como base la lista contenida por defecto en la librería NLTK, que fue complementada con palabras que generan ruido a nivel de negocio y de las geopalabras (nombres de municipios y departamentos).

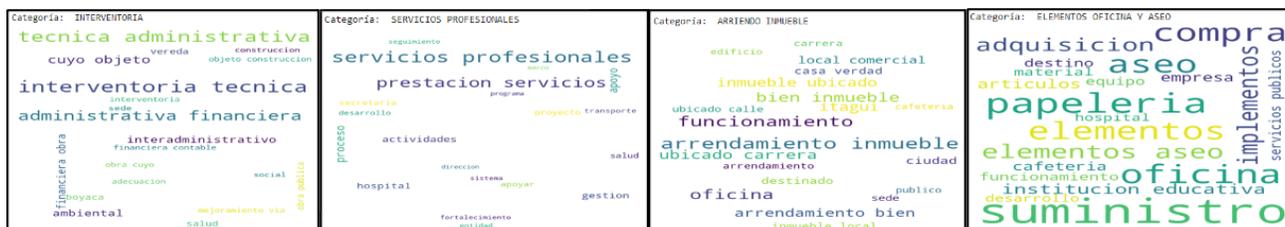
Figura 6. Proceso de preparación de los datos



Fuente: elaboración propia.

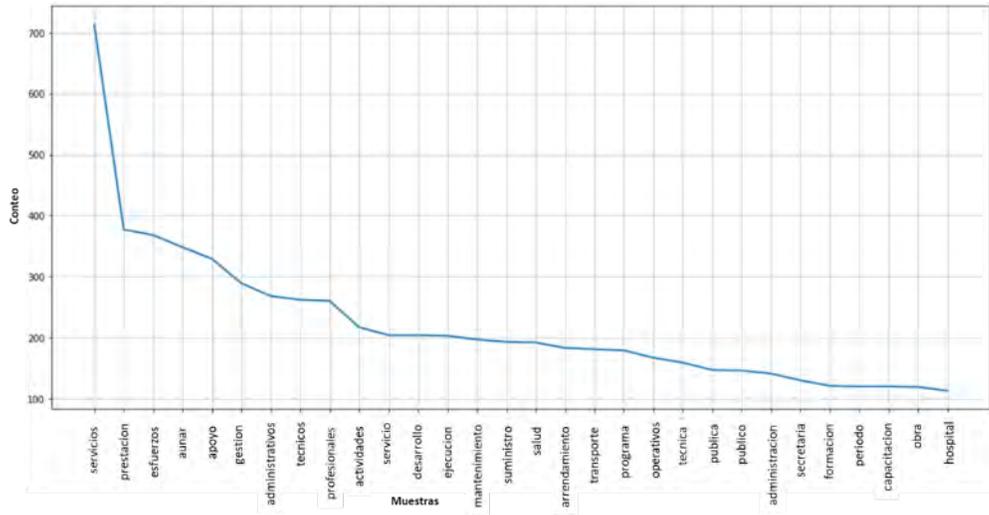
Posteriormente, se realizó el proceso de división de cadenas de texto para determinar, de acuerdo con la frecuencia de aparición en el conjunto de datos, su relevancia dentro del corpus. Así mismo, se construyó una nube de palabras por cada categoría para determinar los términos más representativos dentro de dichas categorías, como se aprecia en las figuras 7 y 8.

Figura 7. Nube de palabras de algunas categorías



Fuente: elaboración propia en Python.

Figura 8. Distribución de principales palabras en el conjunto de datos

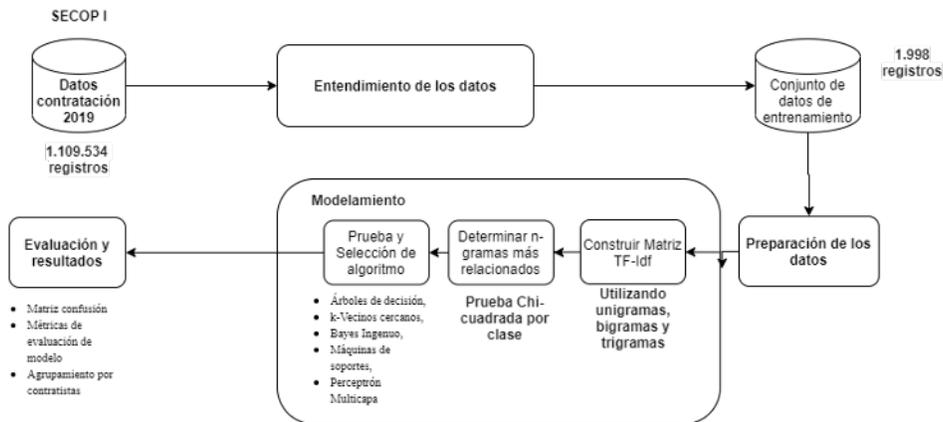


Fuente: elaboración propia en Python.

Modelamiento

En esta fase se construye el modelo a partir de la matriz de frecuencia de términos y se realiza una exploración por medio de los n-gramas seleccionados, para luego realizar una evaluación del comportamiento de los algoritmos y así identificar aquel con la mejor métrica de generalización (figura 9).

Figura 9. Proceso de modelamiento



Fuente: elaboración propia.

Debido a que los algoritmos de aprendizaje y los clasificadores no pueden procesar los textos directamente, se construye una matriz de frecuencia de términos-frecuencia inversa de documentos (TF-IDF) con el vector de cada uno de los objetos contractuales. Para ello se establecieron los siguientes parámetros:

- `sublinear_df = True`, para utilizarlo de forma logarítmica.
- `min_df = 6`, número mínimo de documentos que debe contener una palabra para conservarla.
- `norm = l2`, para utilizarlo con forma euclidiana 1.

- ngram_range = (1,3), puesto que vamos a utilizar, unigramas, bigramas y trigramas.
- stop_words = quitar palabras que contienen, como se indicó en la fase anterior, las palabras en español de la librería NLTK y las geopalabras en Colombia.

El resultado fue una matriz [1.998, 1.750], es decir, 1.998 registros de objetos contractuales con 1.750 características que representan la puntuación para los unigramas, bigramas y trigramas.

Para revisar cuáles son los términos más relacionados en cada categoría por unigramas, bigramas y trigramas, se realizó una prueba chi-cuadrada. En la figura 10 se muestran algunos resultados de esta verificación.

Figura 10: Resultados prueba Chi-cuadrado para los términos más relevantes

```

*'APOYO JURIDICO':
. unigramas más correlacionados:
. abogado
. juridica
. juridico
. bigramas más correlacionados:
. asesoria juridica
. profesionales abogado
. apoyo juridico
. trigramas más correlacionados:
. apoyo juridico procesos
. prestacion servicios profesionales
. servicios profesionales abogado
*'ARRIENDO INMUEBLE':
. unigramas más correlacionados:
. arrendamiento
. ubicado
. inmueble
. bigramas más correlacionados:
. ubicado carrera
. arrendamiento bien
. arrendamiento inmueble
. trigramas más correlacionados:
. inmueble local comercial
. arrendamiento casa verdad
. arrendamiento bien inmueble
    
```

Fuente: elaboración propia en Python.

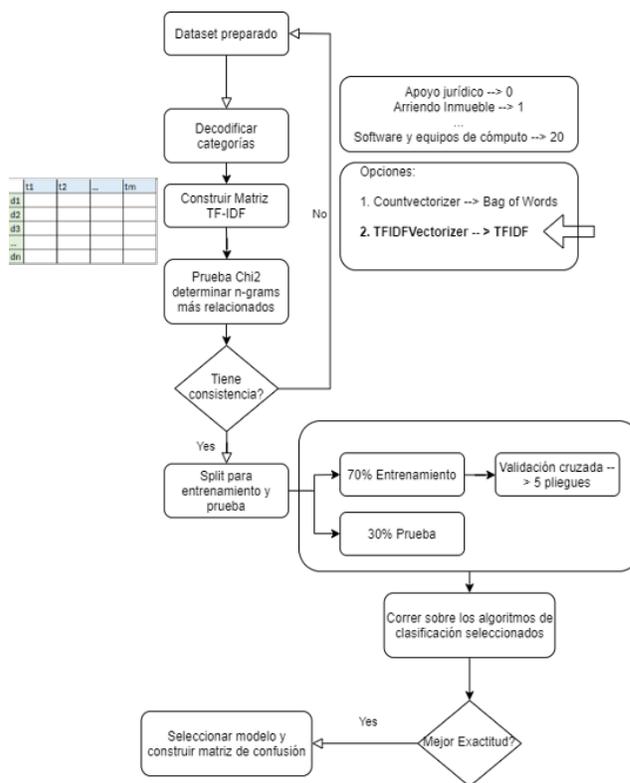
Posteriormente, se realizó la partición del conjunto de entrenamiento y pruebas, el cual se definió en un 30 % a través del método estratificado para balancear las 3 clases con menos registros: “Convenio Gratuidad Educativa”, “Interventoría” y “Obra civil”.

Una vez se tienen las representaciones de los objetos en vectores, estos son utilizados en los modelos

seleccionados para este proyecto. La estrategia utilizada para garantizar la independencia de los datos de entrenamiento y testeo fue utilizar validación cruzada a 5 pliegues, que además permitió aumentar la robustez del estimador y mantener la rigurosidad metodológica en la construcción del modelo. La construcción del proceso descrito del modelamiento se resume en la figura 11.

Con este modelo se obtendrá un avance en la identificación contextual de los objetos contractuales para las métricas en el sector educación, principalmente, que además puede ser aplicable a otros sectores de interés al extender el conjunto de categorías.

Figura 11. Construcción del modelamiento del proyecto



Fuente: elaboración propia.

RESULTADOS

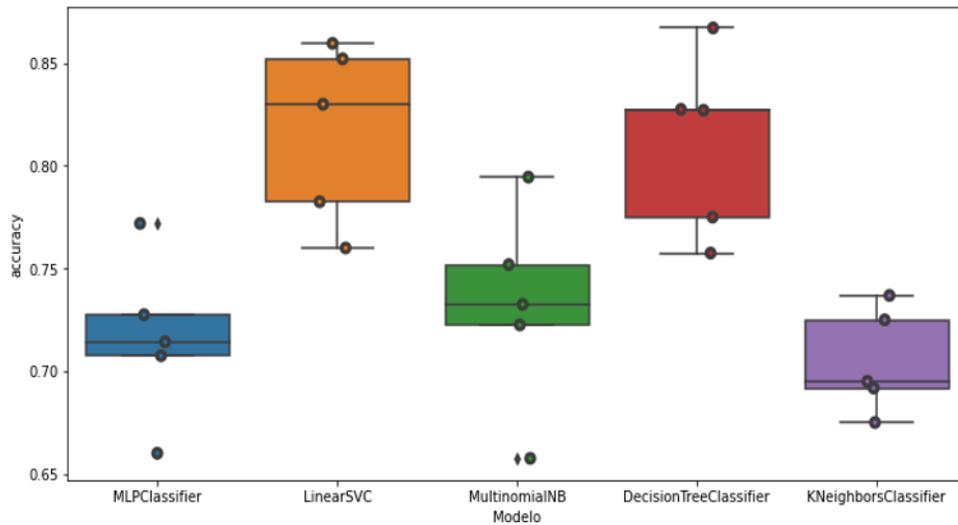
Los hiperparámetros seleccionados para los modelos comparados fueron los siguientes:

- MLPClassifier (hidden_layer_sizes=(80)) → Se establecieron 80 neuronas en 1 capa oculta.

- LinearSVC (random_state=0) → Se conservaron los parámetros establecidos por defecto.²
- MultinomialNB () → Se conservaron los parámetros por defecto.
- DecisionTreeClassifier (random_state=0) → Se conservaron los parámetros por defecto.
- KNeighborsClassifier (n_neighbors=100) → Se conservaron 100 vecinos cercanos.

Una vez seleccionados los modelos, se realizó una comparación para determinar cuál de ellos tiene mejor comportamiento por la métrica de Exactitud (figura 12), permitiendo establecer el siguiente orden: máquina de soporte lineal (81,68 %), árbol de decisión (81,08 %) y perceptrón multicapa (73,28 %), como se muestra en la tabla 3.

Figura 12. Exactitud de los modelos utilizados en el proyecto



Fuente: elaboración propia en Python.

Tabla 3. Porcentajes de exactitud por modelo

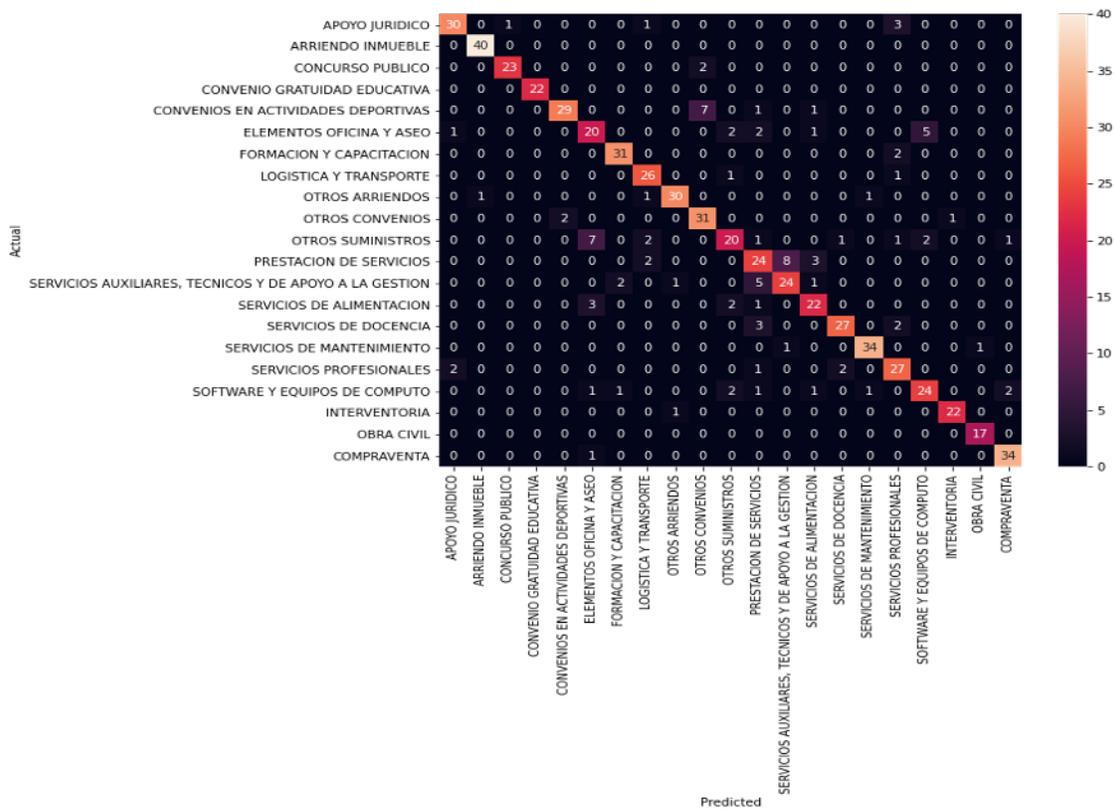
Clasificador	Exactitud (%)
Máquinas de soporte vectorial tipo lineal	81,68
Árboles de decisión	81,08
Clasificador perceptrón multicapa	73,28
Bayes ingenuo multinomial	73,18
K-vecinos cercanos	70,47

Fuente: elaboración propia.

Se seleccionó la máquina de soporte vectorial tipo lineal debido a que este tiene los mejores resultados. Al realizar la prueba del modelo respecto al y-prueba/y-predicción, se obtuvo un comportamiento adecuado en la matriz de confusión (figura 13), señalando que su nivel de generalización es adecuado, sin presentar sobreajuste. Sin embargo, se presentan algunos casos de falsos positivos en la clasificación para las categorías “Servicios auxiliares de apoyo a la gestión”, “elementos de oficina y aseo” y “otros convenios”, principalmente.

² Los parámetros por defecto, son aquellos con los cuales vienen predeterminados cada algoritmo en Python.

Figura 13. Matriz de confusión de valores predichos vs. valores de prueba



Fuente: elaboración propia en Python.

En el análisis de los casos que en la categoría “convenios de actividades deportivas” fueron erróneamente clasificados como “otros convenios”, se observa que el contenido gramatical sí pertenece al conjunto de convenios, por lo que este nivel de error puede ser aceptable en el contexto del negocio, así como los contratos de prestación de servicios que fueron pronosticados como “servicios auxiliares, técnicos y de apoyo a la gestión”, lo cual se debe a que esta es una categoría que sigue estando en la misma línea categórica. Para el caso de los registros que siendo de “otros suministros” fueron pronosticados como “Elementos

de Oficina y Aseo”, se podría considerar aumentar el tamaño de muestra para que el corpus de la matriz “documentos términos” contemple vocablos que permitan clasificar mejor esta categoría en una siguiente iteración. Finalmente, se realizó una exploración de los términos más relacionados mediante una prueba chi-cuadrada, encontrando que los términos se clasificaron acordes con las categorías predichas.

La tabla 4 muestra las categorías con el resultado del modelo en las métricas de Puntuación-F1, Sensibilidad y Precisión.

Tabla 4. Métricas de comportamiento del modelo

Categorías	Precisión	Sensibilidad	Puntuación F1	Casos
Apoyo jurídico	0,91	0,86	0,88	35
Arriendo inmueble	0,98	1,00	0,99	40
Concurso público	0,96	0,92	0,94	25
Convenio gratuidad educativa	1,00	1,00	1,00	22
Convenios en actividades deportivas	0,94	0,76	0,84	38
Elementos oficina y aseo	0,62	0,65	0,63	31

Formación y capacitación	0,91	0,94	0,93	33
Logística y transporte	0,81	0,93	0,87	28
Otros arriendos	0,94	0,91	0,92	33
Otros convenios	0,78	0,91	0,84	34
Otros suministros	0,74	0,57	0,65	35
Prestación de servicios	0,62	0,65	0,63	37
Servicios auxiliares, técnicos y de apoyo a la gestión	0,73	0,73	0,73	33
Servicios de alimentación	0,76	0,79	0,77	28
Servicios de docencia	0,90	0,84	0,87	32
Servicios de mantenimiento	0,94	0,94	0,94	36
Servicios profesionales	0,75	0,84	0,79	32
Software y equipos de cómputo	0,77	0,73	0,75	33
Interventoría	0,96	0,96	0,96	23
Obra civil	0,94	1,00	0,97	17
Compraventa	0,92	0,97	0,94	35
Exactitud			0,84	660
Promedio macro	0,85	0,85	0,85	660
Promedio ponderado	0,85	0,84	0,84	660

Fuente: elaboración propia en Python.

Las clases con mejor Puntuación-F1 son “CONVENIO Y GRATUIDAD EDUCATIVA” (100 %), “ARRIENDO INMUEBLE” (99 %), “OBRA CIVIL” (97 %) e “INTERVENTORÍA” (96 %). Es muy probable que esto se deba a que dichas categorías son bastante específicas dentro del contexto del contrato y, por ende, se generalizan adecuadamente. Por el contrario, las clases con menor Puntuación-F1 son “ELEMENTOS DE OFICINA Y ASEO” (63 %), “PRESTACIÓN DE SERVICIOS” (63 %) y “OTROS SUMINISTROS” (65 %), puesto que, como se expuso previamente, estas pueden crear ambigüedades con otras categorías que se encuentren en un contexto similar. La exactitud general para el modelo lineal con los datos de predicción es de 84 %, lo cual es bastante bueno para el propósito del proyecto.

Como el objetivo final es identificar los contratistas multiobjeto, o aquellos que están presentes en más de dos categorías, se realiza un agrupamiento por identificador de contratista y el número de categorías que se les asignó de acuerdo al clasificador construido. Se realizó el conteo bajo estos parámetros para el conjunto de datos de los 10.120 registros del sector educación, como se aprecia en la tabla 5.

Tabla 5. Conteo de cantidad de categorías por identificador del contratista

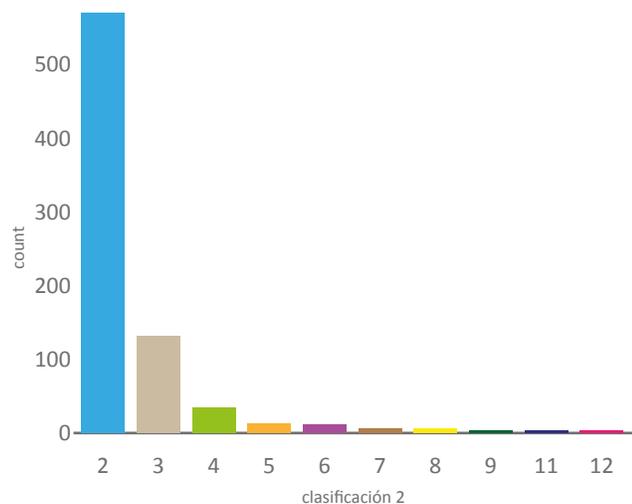
	Identificación del Contratista	Clasificación2
0	899999063	12
1	900761702	11
2	900568326	9
3	30284542	8
4	800157163	8
5	860070374	8
6	900600059	8
7	901104562	7
8	900176990	7
9	901274854	7
10	810004774	7

Fuente: elaboración propia en Python.

Por último, se realiza un histograma (figura 14) para comprobar la distribución de los contratistas que tienen dos o más categorías, con el fin de que el proceso auditor pueda focalizar sus esfuerzos. Es evidente que

este agrupamiento permitirá revisar los contratos sobre aquellos contratistas que tienen 3 o más categorías y así revisar su correcta ejecución.

Figura 14. Métricas de comportamiento del modelo



Fuente: elaboración propia en Python.

DISCUSIÓN

En el ámbito del aprendizaje automático, es de amplio conocimiento que el éxito de un modelo depende en gran medida de la calidad de los datos de entrada, más aún en el caso de aquellos enfocados a minería de texto para clasificación. Para el presente caso, el conjunto de datos de contratos en la variable de entrada, que es el objeto contractual, presenta inconvenientes de ambigüedad respecto al contexto del fin contractual; es decir, la finalidad del objeto a contratar puede confundir al algoritmo debido a la redacción del mismo, llevando a clasificarlo equivocadamente en otro (por ejemplo, “SERVICIOS PRESTAR LOS PROFESIONALES PARA BRINDAR APOYO LOGÍSTICO EN LAS OLIMPIADAS DEPORTIVAS DEL MUNICIPIO XYZ”, en el cual se perciben 3 categorías: prestación de servicios, logística y transporte y convenios en actividades deportivas).

Otro aspecto que se detectó en la construcción de este modelo, y que puede mejorarse en futuras iteraciones, es la cantidad de información respecto a la descripción del objeto, que en algunos casos puede

ser insuficiente para que el modelo tome una decisión acertada al momento de clasificar (por ejemplo, objeto = “MÉDICO GENERAL”, “ENFERMERA”).

Por otra parte, el volumen de datos por categoría hace parte esencial de este tipo de modelos, por lo cual se recomienda tener la mayor cantidad de registros preclasificados por categoría, y que estos se encuentren balanceados, pues a pesar de utilizar la estrategia de estratificación dentro de los parámetros de particionamiento en entrenamiento y prueba, es preferible que los datos se encuentren balanceados desde el inicio para reducir el nivel de sesgo en la clasificación.

En general, es necesario crear categorías específicas para cada sector, salvo aquellas que sean transversales (como apoyo jurídico, entre otros). Para ello, se recomienda que el ajuste se lleve a cabo tras un análisis preliminar de los objetos y los contextos con un grupo de expertos del negocio, para así determinar de manera focalizada las categorías óptimas del sector estudiado.

Es posible mejorar el comportamiento y la capacidad de generalización de los modelos mediante la exploración de parámetros optimizados, lo cual probablemente no solo permitirá aumentar una métrica como la eficacia, sino incluso marcar la diferencia de comportamientos entre los dos modelos que mayor exactitud obtuvieron (máquina de soporte vectorial y árbol de decisión).

Respecto al comportamiento de n-gramas, los resultados muestran que son los unigramas y bigramas los que aportan más a la generalización del modelo seleccionado.

Finalmente, en el agrupamiento de los casos de presuntos multiobjeto se evidencian también clasificaciones de índole similar en la clase, situación que puede aumentar la presencia de falsos positivos. En una próxima iteración, estos últimos pueden mejorar la asertividad de la selección de la muestra mediante reglas de asignación de scores de incompatibilidad entre clases para que el auditor revise aquellos contratistas que además de tener más de 2 clases en su conteo tengan presencia de clases incompatibles en contratos suscritos.

CONCLUSIONES

Ampliar el número de las 6 categorías inicialmente seleccionadas mejoró el desempeño del algoritmo debido a la ambigüedad en el etiquetado manual realizado por los anotadores expertos del negocio para los conceptos de prestación de servicios y suministros, lo que ocasionaba que el conjunto de datos para entrenamiento no fuera lo suficientemente adecuado. Lo anterior se subsanó mediante la creación de 21 subcategorías específicas a partir de las 6 iniciales, que los anotadores etiquetaron de una manera más precisa, pues ya no se creaba la ambigüedad en las dos categorías que lo generaban.

Los algoritmos de máquinas de soporte vectorial tipo lineal fueron los de mayor desempeño al momento de clasificar las categorías de los textos de los objetos contractuales. Esto es consistente con el estado del arte sobre el tema, donde se considera que es uno de los mejores algoritmos para este tipo de ejercicios.

El hecho de que un contratista se encuentre en dos o más categorías permitiría la revisión del asunto por parte del proceso auditor de la Contraloría General de la República, permitiendo identificar un posible contratista multiobjeto y además analizar comportamientos de posible cartelización. Sin embargo, vale la pena mencionar que se realizó un ejercicio exploratorio con el primer contratista que obtuvo presencia en 12 de las 21 categorías, el cual resultó ser una institución universitaria que dentro de su actividad misional desarrolla más de una actividad en diferentes ramas (estudios, consultoría, entre otros). Al respecto, resulta pertinente aclarar que en otro tipo de organización esto no sería un hecho común, por lo que resulta conveniente revisar este tipo de casos con mayor detalle.

REFERENCIAS

- Al-Amini, H. S. (2020). The future of public sector auditing: Living in times of change. *International Journal of Government Auditing*, 47(1), 4-5. http://intosaijournal.org/wp-content/uploads/2020/02/INTOSAI-Journal_Winter-2020.pdf
- Álvarez-Jareño, J. A., Badal-Valero, E., & Pavía, J. M. (2018). Aplicación de métodos estadísticos, económicos y de aprendizaje automático para la detección de la corrupción. *Revista Internacional de Transparencia e Integridad*, 9, 1-11. <https://dialnet.unirioja.es/servlet/articulo?codigo=6977094>
- Bologa, A. R., Bologa, R., & Flores, A. (2010). Big data and specific analysis methods for insurance fraud detection. *Database Systems Journal*, 1(1), 30-39.
- Contraloría General de la República de Colombia [CGR]. (2018). *Plan Estratégico CGR 2018-2022*. CGR.
- Córdoba-Larrarte, C. F. (2019). Océano: monitoreo eficiente en la contratación pública. *Economía Colombiana*, 356, 4-5. <https://www.economicolombiana.co/revista/oceano-393>
- García, J., Molina, J. M., Berlanga, A., Patricio, M. A., Bustamante, A. L., & Padilla, W. R. (2018). *Ciencia de datos. Técnicas analíticas y aprendizaje estadístico*. Alfaomega.
- Giraldo-Polanía, L. A., Parra-Ortiz, J. W., Cotrino-García, Y., Dulce-Vanegas, M. F., & Tafur-Díaz, J. (2018). *Big data. Análisis de caso en la Contraloría de Bogotá con la entrega de bonos*. Contraloría de Bogotá. <https://www.olacefs.com/wp-content/uploads/2018/10/1%c2%b0-Premio-Bogot%c3%a1-Colombia.pdf>
- Hsu, B. M. (2020). Comparison of supervised classification models on textual data. *Mathematics*, 8(5). <https://doi.org/10.3390/MATH8050851>
- Li, S. (2018). *Multi-class text classification with scikit-learn*. <https://towardsdatascience.com/multi-class-text-classification-with-scikit-learn-12f1e60e0a9f>

Mohamed, A. (2005). *Survey on multiclass classification methods*. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.175.107&rep=rep1&type=pdf>

Organización Internacional de Entidades Fiscalizadoras Superiores [INTOSAI] (s.f.) *About us*. <https://www.intosai.org/about-us>

Othman, R., Aris, N. A., Mardiyah, A., Zainan, N., & Amin, N. M. (2015). Fraud detection and prevention methods in the Malaysian public sector: Accountants' and internal auditors' perceptions. *Procedia Economics and Finance*, 28(April), 59-67. [https://doi.org/10.1016/s2212-5671\(15\)01082-5](https://doi.org/10.1016/s2212-5671(15)01082-5)

Pedregosa, F., Varoquaux, G., Gramfor, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Pettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.

Rennie, J. D. M., & Rifkin, R. (2001). *Improving multiclass text classification with the support vector machine*. Massachusetts Institute of Technology.

Song, Y. Y., & Lu, Y. (2015). Decision tree methods: Applications for classification and prediction. *Shanghai Archives of Psychiatry*, 27(2), 130-135. <https://doi.org/10.11919/j.issn.1002-0829.2>

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*. [http://citeseerx.ist.psu.edu/viewdoc/summary?](http://citeseerx.ist.psu.edu/viewdoc/summary?https://doi.org/10.1.1.198.5133) <https://doi.org/10.1.1.198.5133>

Predicción de la serie temporal del indicador bancario de referencia (IBR) con redes neuronales

Predicting the Benchmark Banking Indicator (IBR) Time Series through Neural Networks

Germán Enrique Coy Mondragón^{ab}, Óscar Granados^{ac}, Olmer Garcia-Bedoya^{ad}

^a Universidad de Bogotá Jorge Tadeo Lozano, Colombia

^b germane.coymondragon@utadeo.edu.co

^c <https://orcid.org/0000-0002-4992-8972>

^d <https://orcid.org/0000-0002-6964-3034>

RESUMEN

En años recientes, la predicción del comportamiento del Indicador Bancario de Referencia (IBR) se ha hecho relevante debido a su importancia en el mercado monetario colombiano. El propósito de este trabajo es demostrar la eficiencia de las redes LSTM en la generación de predicciones de series de tiempo —a través de su memoria a corto y largo plazo— que sean comparables con el modelo predictivo ARIMA para estudios econométricos. Se analizó la incidencia de la tasa representativa de mercado (TRM) y la tasa de los bonos de la deuda pública (TES) a 10 años, comparando ambos indicadores con el IBR. Con lo anterior, se buscó determinar la correlación existente entre estas variables mediante el método de Pearson. Finalmente, la eficiencia del modelo fue evaluada con el error cuadrático medio (RMSE), utilizando una red LSTM multivariable con tres entradas (IBR, TES y TRM) y una salida.

Palabras clave: series de tiempo, aprendizaje automático, redes LSTM, ARIMA, coeficiente de correlación de Pearson.

Citation: Coy-Mondragón, G. E., Granados, O. y Garcia-Bedoya, O. (2021).

Predicción de la serie temporal del indicador bancario de referencia (IBR) con redes neuronales. *Mutis*, 11(1), 65-76.

<https://doi.org/10.21789/22561498.1748>

Recibido: Enero 18, 2021. **Aceptado:** Marzo 30, 2021.

Copyright: ©2021. Chávez-Plazas, Y. A., Ramírez-Mahecha, M. C. y Camacho-Kurmen, J. E. (2021). This is an open-access article, which permits unrestricted use, distributions and reproduction in any medium, provided the original author and source are credited.

Competing Interests: The authors have no conflict of interest.

ABSTRACT

In recent years, predicting the behavior of the Benchmark Banking Indicator (IBR) has become relevant due to its importance in the Colombian money market. The purpose of this paper is to demonstrate the efficiency of LSTM networks for generating predictions of time series —through their long and short-term memory— that are comparable with the ARIMA predictive model for econometric studies. The incidence of the representative market rate (TRM) and

the rate of 10-year public debt bonds (TES) was analyzed and compared to the IBR, seeking to determine its correlation through the Pearson method. Finally, the model efficiency was evaluated with the mean square error (RMSE), using a multivariable LSTM network with three inputs (IBR, TES, and TRM) and one output.

Keywords: Time series, machine learning, LSTM network, Root Mean Square Error (RMSE), ARIMA, Pearson Correlation Coefficient.

INTRODUCCIÓN

El Banco de la República y el Ministerio de Hacienda y Crédito Público de Colombia, con el propósito de tener un indicador confiable que reflejara la liquidez del mercado monetario, crearon en 2008 el Indicador Bancario de Referencia (IBR). Este indicador relaciona el precio al que los bancos están dispuestos a ofrecer o captar recursos en el mercado interbancario, lo que convierte al IBR en una referencia indispensable para las operaciones financieras de corto plazo. En este mecanismo se establecen cuatro plazos de cotización (*overnight*, un mes, tres meses y seis meses) y a partir de él se calcula la mediana de las cotizaciones de los participantes que se publica en los sistemas de información del Banco de la República. Varios mercados utilizan indicadores semejantes que se convirtieron en referentes para el IBR, principalmente, el London InterBank Offered Rate (LIBOR) de Londres, el Euro Interbank Offered Rate (EURIBOR) de la zona Euro y la tasa de interés interbancaria de equilibrio (TIIE) de México, por listar algunos (Abrantes-Metz *et al.*, 2001; Duffie & Stein, 2015; Jamshidian, 1997). Estos indicadores utilizan las tasas de referencia interbancaria, por lo que varios estudios han buscado determinar su pronóstico a través de técnicas de aprendizaje automático y, más recientemente, de aprendizaje profundo.

En Colombia se registran pocos trabajos similares con esta clase de técnicas. Por ello, entendiendo la relevancia de estos indicadores para los mercados financieros, surge la necesidad de desarrollar un modelo de predicción del comportamiento del IBR. Diversos estudios han utilizado modelos de series temporales, con la ayuda de los principios del aprendizaje automático y el uso de técnicas como la red LSTM, para hallar la predicción de la tendencia en diferentes casos de estudio, arrojando importantes resultados. En los últimos años, se ha utilizado la metodología LSTM para predecir y clasificar texto con la ayuda de redes neuronales recurrentes (RNN), entre ellas, las redes LSTM (Liu *et al.*, 2016), así como para predecir las

cargas de trabajo (Kumar *et al.*, 2018), la dependencia a largo plazo de series temporales con datos financieros (Zhang *et al.*, 2019), la detección temprana de un shock séptico con el objetivo de reducir la mortalidad en pacientes (Fagerström *et al.*, 2019), los precios de las acciones en la apertura del mercado bursátil (Qui *et al.*, 2020), la trayectoria de un vehículo con datos de condición reales (Bükk & Johansson, 2020) o los lugares para la exploración minera (Li *et al.*, 2020), por citar algunos ejemplos.

Algunos estudios usan las metodologías de aprendizaje automático para predecir indicadores financieros y económicos, empleando el algoritmo genético puro (GA) y el algoritmo genético integrado con la red neural recurrente (GA + RNN) para predecir la tasa LIBOR en USD (Tan, 2019). Así mismo, existen propuestas que buscan predecir a través de redes neuronales el cambio diario del precio de la tasa de interés LIBOR y el índice bursátil estadounidense Standard and Poor's 500 (S&P 500) (Dmitrieva *et al.*, 2002), o pronosticar las acciones de mayor bursatilidad en Colombia (Ecopetrol y Preferencial Bancolombia) que se encuentran cotizadas en el mercado local y en el estadounidense, a través del esquema de American Depositary Receipt (ADR) (Villada *et al.*, 2012), por ejemplo. Otros trabajos han usado esta metodología para pronosticar variables económicas como la inflación, para el caso de México, a través de una red neural artificial (RNA) con un enfoque estacional y un proceso AR (Robles, 2019); en Costa Rica, al comparar el rendimiento de algunas técnicas de *machine learning* (K-Nearest Neighbours [KNN], bosques aleatorios, *extreme gradient boosting* y un modelo de tipo *long short-term memory* [LSTM]) (Rodríguez-Vargas, 2020); o en Chile para pronosticar la inflación subyacente del IPC con métodos de *machine learning* y *big data* (Leal *et al.*, 2020). Finalmente, existen algunos trabajos que hacen análisis mediante técnicas de *machine learning* como las redes LSTM y ARIMA (*autoregressive integrated moving average*) para predecir el precio de cierre de las criptomonedas (Castillo-Mur, 2020).

Teniendo en cuenta lo anterior, ¿cómo podría aplicarse esta metodología a un indicador interbancario?, ¿cómo integrarla en la interacción con otras variables financieras? A partir de estos interrogantes, la presente investigación busca contribuir en la definición y generación de una predicción confiable de la serie temporal del IBR y analizar en un ambiente comparativo la eficiencia de redes LSTM tanto supervisadas como no supervisadas. Adicionalmente, esta investigación busca establecer la confiabilidad y las limitantes del uso de herramientas de aprendizaje automático en variables financieras en las que habitualmente los análisis econométricos han sido ampliamente usados.

Para lograr los objetivos propuestos, esta trabajo usa la metodología CRISP-DM, la cual permite realizar proyectos mediante una estructura de tareas por nivel o procesos (Wirth *et al.*, 2000). Esta metodología se fundamenta, primero, en la comprensión del negocio —definida en esta introducción— y se apoya en la sección de preliminares. Además, se apoya en la comprensión y preparación de los datos que se desarrollan en la metodología, junto con el modelo y su implementación. La evaluación de los resultados se integra en la sección de resultados. Finalmente, la evaluación de los resultados y la identificación de la utilidad del modelo se presenta en la sección de discusión.

PRELIMINARES

En esta sección se hace una revisión básica de algunos conceptos que son empleados en el desarrollo del documento y que facilitan el entendimiento de su integración.

Indicador bancario de referencia (IBR)

La investigación se desarrolló con el esquema *overnight* del IBR, el cual se comporta de la siguiente manera, de acuerdo con Cardozo y Murcia (2014):

- Las cotizaciones por parte de las entidades financieras participantes se realizan diariamente entre 10:00 a. m. y 10:15 a. m. Luego, el Banco de la República toma en cuenta solo la última tasa de cotización recibida de cada entidad participante.
- El Banco de la República publica, cada día hábil a las 11:00 a. m., el IBR *overnight*.

- Las transferencias de los recursos interbancarios se deberán realizar antes del cierre de la plataforma de Servicios Electrónicos del Banco de la República (SEBRA).
- El pago o la devolución de los créditos interbancarios entre los participantes se realiza de manera automática a través de la plataforma que el Banco de la República disponga para el efecto.

Coefficiente de correlación

El coeficiente de correlación busca establecer si existe alguna asociación entre dos variables cuantitativas aleatorias que tienden a ser mayores (o menores) a medida que aumentan los valores de la otra variable; su correlación puede determinarse mediante el resultado, que puede fluctuar entre 1 y -1, mientras que su proximidad a 0 puede indicar la no correlación o baja correlación lineal (ec. 1).

$$-1 \leq \rho \leq 1 \quad (\text{Ec. 1})$$

Coefficiente de correlación Pearson

Con el coeficiente de correlación Pearson se puede observar las similitudes ρ entre las variables o los individuos. Así mismo, es posible medir la fuerza o el grado de asociación que posee una distribución normal bivariada conjunta (Restrepo & González, 2007). Este coeficiente se expresa de la siguiente manera (ec. 2):

$$\rho = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} \quad (\text{Ec. 2})$$

donde:

- si ρ es positivo la relación es directa entre las variables;
- si ρ es negativo la relación entre las variables es inversa, es decir, cuando una aumenta, la otra disminuye;
- si ρ es mayor de 1 o -1 hay una relación perfecta entre ambas variables;
- si ρ es igual a 0 no hay relación entre las variables; sin embargo, puede existir una relación no lineal.

ARIMA

El modelo ARIMA (*autoregressive integrated moving average*) es un proceso estocástico para el estudio de series temporales en temas econométricos desarrollado por Box y Jenkins. Este modelo se deriva de tres componentes —AR (autorregresivo), I (integrado) y MA (medidas móviles)— y logra interpretar valores en funciones lineales con componentes cíclicos o estacionales. Además, este supone valores futuros combinando valores pasados, si estos son lineales (Pai & Lin, 2005), como se expresa en la ecuación 3. Sin embargo, si el grupo de datos no corresponden a una linealidad o no se adecua a este comportamiento, el modelo ARIMA se ve limitado.

$$\rho = y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (\text{Ec. 3})$$

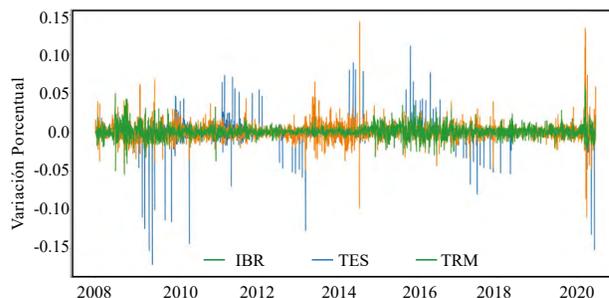
MATERIALES Y MÉTODOS

Preparación de los datos

Los datos de entrada están constituidos por el IBR, los títulos de deuda pública a 10 años (TES 10 años) y la tasa representativa del mercado (TRM) durante el período comprendido entre el 2 de enero de 2008 y el 15 de junio de 2020. Los datos se obtuvieron del conjunto de datos públicos del Banco de la República de Colombia y fueron ajustados para ser cargados en Python.

El conjunto de datos del plazo *overnight* del IBR está dividido en 21 columnas (información proporcionada por los bancos participantes para el cálculo del indicador y el cálculo del IBR nominal y efectivo) y se define como un valor relativo. Los datos con los que se comparó el indicador fueron la TRM, expresada en pesos colombianos (COP) por dólar estadounidense (USD), y los rendimientos de los títulos de deuda pública a diez años (TES), expresados en la tasa de rendimiento, que es un valor relativo. Estos últimos se tomaron en el mismo rango de tiempo y, al igual que el IBR, fueron descargados en formato csv, siendo necesario mejorar la calidad de los datos antes de empezar su análisis con el modelo de aprendizaje automático. Para lograr datos óptimos, se determinó la tasa de variación de cada uno de los conjuntos de datos comparados con el fin de obtener datos semejantes y así poder analizar de forma más clara los cambios que experimentan los datos a través del tiempo (figura 1).

Figura 1. Tasa de variación



Fuente: elaboración propia.

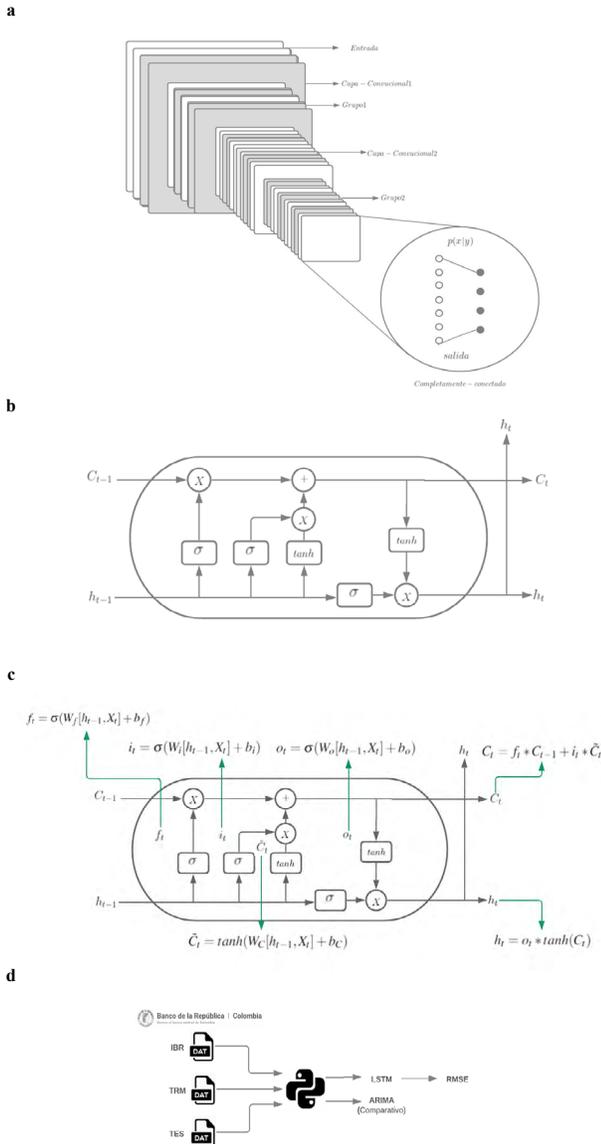
Modelo

El método para la predicción del IBR está constituido por una red neuronal recurrente LSTM (figura 2a) que aprende de los datos históricos a largo plazo. Este método, utilizado en aprendizaje profundo (*deep learning*), aprende los valores futuros de una serie temporal en función de los valores anteriores y, como toda RNN, permite conexiones arbitrarias entre las neuronas, generando temporalidad y, desde luego, memoria.

La red LSTM tiene una arquitectura recurrente con conexiones que retroalimentan las capas de la red, permitiendo analizar los mismos problemas que una máquina de Turing. Una red neuronal recurrente tradicional está compuesta por una sola capa que calcula el peso que van a tener los datos históricos. En las redes LSTM del sistema de persistencia se definen cuatro puertas, o unidades, que interactúan entre sí (figura 2b). Por lo tanto, las redes LSTM solucionan el problema de largo plazo con el que se encuentran las redes neuronales recurrentes tradicionales, logrando resolver el

problema que surge cuando se necesitan datos de muchos ciclos atrás para realizar la predicción del ciclo siguiente (Pastor-Cobo & Torre-Barrio, 2019).

Figura 2. Modelo LSTM



a. Estructura LSTM implementada; b. Primer modelo LSTM; c. Segundo modelo LSTM; d. Proceso del modelo.

Fuente: elaboración propia.

Este algoritmo LSTM utiliza una de las formas de RNN más comunes y es capaz de procesar una secuencia de longitudes arbitrarias, aplicando recursivamente una función de transición a su vector de estado oculto

interno en las secuencias de entrada (Liu *et al.*, 2016). Además, este recurso evita problemas de dependencia a largo plazo para la predicción de series de tiempo. Según Hochreiter y Schmidhuber (citados en Qui *et al.*, 2020), este modelo consiste en un conjunto único de células de memoria que reemplazan las neuronas de la capa oculta del RNN, y su clave es el estado de las celdas de memoria que filtran información a través de la estructura de la puerta. Matemáticamente, una celda de la red LSTM se representa de la siguiente manera (figura 2c).

- Entrada

X_t = valor de la secuencia en el instante t

h_{t-1} = salida de la unidad LSTM en el paso anterior

C_{t-1} = estado de la unidad LSTM

- Salida

Se conforma por un nuevo estado o una neurona autorrecurrente, que es la combinación entre C_{t-1} y el candidato a nuevo estado \tilde{C}_{t-1} , donde podemos ver f_t y i_t como el grado de influencia de C_{t-1} y \tilde{C}_t .

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (\text{Ec. 4})$$

La salida de la unidad o celda sería:

$$h_t = o_t * \tanh(C_t) \quad (\text{Ec. 5})$$

Para hallar el nuevo candidato se utiliza la siguiente ecuación:

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, X_t] + b_C) \quad (\text{Ec. 6})$$

donde $W_C[h_{t-1}, X_t]$ representa las entradas y los datos recurrentes, y $[h_{t-1}, X_t]$ concatena los vectores h_{t-1}, X_t y b_C .

Con el nuevo candidato hallado se calcula un nuevo estado:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (\text{Ec. 7})$$

Ahora, se calcula el f_t *forget gate*, o puerta de olvido,

$$f_t = \sigma(W_f[h_{t-1}, X_t] + b_f) \quad (\text{Ec. 8})$$

y la i_t *external input gate*, o puerta de entrada externa,

$$i_t = \sigma(W_i[h_{t-1}, X_t] + b_i) \quad (\text{Ec. 9})$$

Para terminar, se calcula otra vez el nuevo estado con la ecuación 7, se calcula la salida de la red LSTM con la ecuación 5, y se determina O_t *output gate*, o puerta de salida.

$$o_t = \sigma(W_o[h_{t-1}, X_t] + b_o) \quad (\text{Ec. 10})$$

Finalmente, para validar los datos obtenidos primero se compararán los resultados de la red LSTM con los de ARIMA, a fin de determinar su calidad en la predicción. Seguidamente, para determinar el error del modelo se evalúa el error cuadrático medio (RMSE, en inglés) con una red LSTM multivariable (figura 2d).

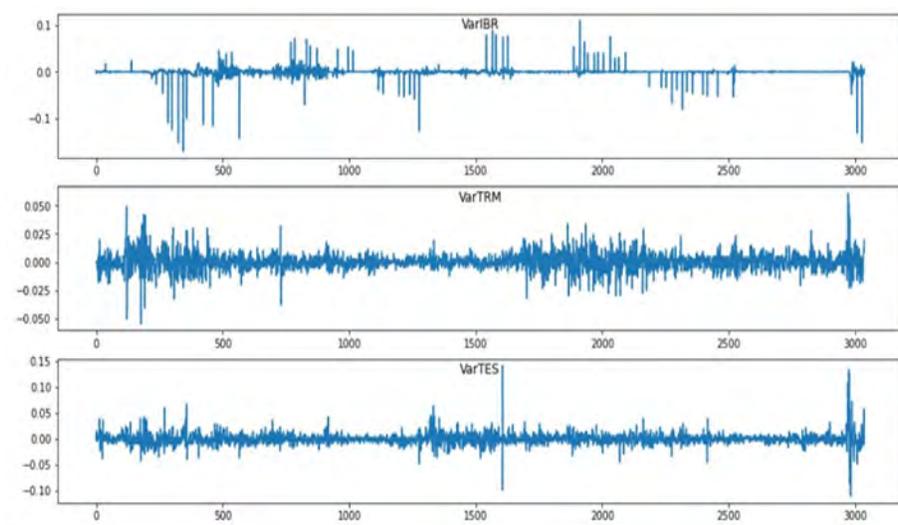
Implementación del modelo

Predicción del IBR, TES y TRM

En esta primera etapa de la investigación se determinó la predicción independiente de cada una de las series temporales definidas en el mismo rango de tiempo (figura 3), haciendo uso de redes LSTM con 50 neuronas en la primera capa oculta y una neurona en la capa de salida. Esta última fue ajustada a 10 épocas con un tamaño de 20 lotes.

Posteriormente, se evaluó la predicción de estas mismas variables con el modelo ARIMA, comparando sus resultados con los hallazgos obtenidos de la predicción con la red LSTM. Finalmente, se calcularon algunos comportamientos estadísticos de las variables y se determinó su correlación.

Figura 3. Señales de los indicadores usados IBR, TES y TRM



Fuente: elaboración propia.

Cálculo del error de la red LSTM

En esta segunda etapa de la investigación, tras comprobar la eficiencia y calidad de la predicción del modelo de aprendizaje automático LSTM frente al ARIMA, se procedió a comprobar la validez del modelo propuesto con ayuda de una red LSTM múltiple con tres entradas (ver figura 7) y una salida, donde se evaluó el modelo con ayuda del RMSE (ecuación 11) y el error absoluto medio (MAE) (ec. 12).

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad (\text{Ec. 11})$$

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad (\text{Ec. 12})$$

Mediante el uso del error cuadrático medio, se evalúa la diferencia entre los datos reales y los pronosticados, calculando la media de los cuadrados del error. De otro lado, con el uso de MAE se puede limitar el error, donde al elevar al cuadrado la diferencia solo se observa los errores más altos, afectando así la calidad del resultado final.

Inicialmente, se normalizó las series temporales, o señales de entrada, y se convirtió el problema a uno de tipo aprendizaje supervisado. La importancia de hacer este cambio radica en la necesidad de contar con patrones de entrada (x) y salida (y) para que el algoritmo pueda aprender a predecir el patrón salida con los patrones de entrada, convirtiendo los DataFrame con $t-1$ como x y t como y . Al ejecutar el código podemos ver las tres series temporales (señales) de entrada y la variable $var1(t)$, que corresponde a la predicción de salida (figura 4).

Figura 4. Conjunto de datos transformados

	var1(t-1)	var2(t-1)	var3(t-1)	var1(t)
1	0.607475	0.489499	0.438874	0.615705
2	0.615705	0.428184	0.500116	0.598145
3	0.598145	0.508808	0.442443	0.612714
4	0.612714	0.517276	0.419458	0.611586
5	0.611586	0.150407	0.464547	0.609342

(3036, 4)

Fuente: elaboración propia.

Posteriormente, se definió y se ajustó el modelo, asignando 50 neuronas para la primera capa oculta y una neurona para la capa de salida, junto con 50 épocas de entrenamiento y un tamaño de 20 lotes. Finalmente, se ejecuta el modelo con aproximadamente 200 horas de datos para entrenamiento y 2.800 horas para pruebas, tanto para las variables de entrada como para las de salida, realizando seguimiento al entrenamiento y a la pérdida de pruebas durante este.

RESULTADOS

Con el fin de identificar la incidencia entre los conjuntos de datos del IBR y su comparación con la TRM y la tasa de rendimiento de los TES a 10 años, se encontraron los siguientes resultados.

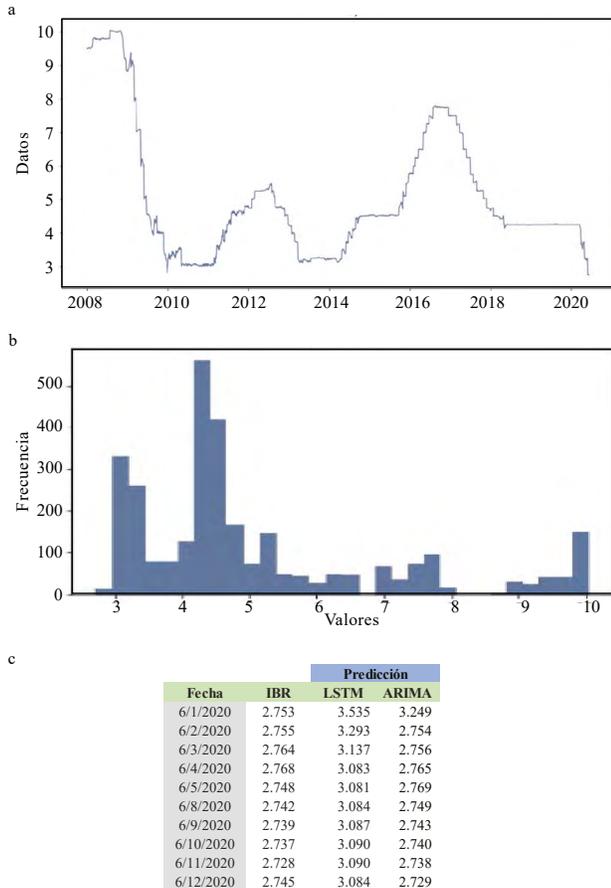
Análisis del IBR

El IBR surgió como respuesta a las anomalías crediticias desencadenadas por el DTF y con el objetivo de elevar estándares de credibilidad y transparencia. Desde sus inicios, en 2008, este indicador presentó algunas fluctuaciones crecientes y decrecientes, siendo las más significativas aquellas ocurridas en los periodos 2008-2010 y 2016-2018. En 2020, desde marzo, se evidencia que este indicador decreció hasta 2,742 %, como resultado del impacto de la pandemia por COVID-19 (figura 5a).

Según la frecuencia de los datos, se determinó que la mayoría de estos se encuentran en un rango de 4-5 % y fluctúa entre 2,728 y 10,044 %, con una media = 5,0661, mediana = 4,499, desviación estándar = 1,9098 y varianza = 3,6474 (figura 5b). Adicionalmente, se comparó la eficiencia de los modelos predictivos Red LSTM con los modelos ARIMA (figura 5c). De acuerdo con los datos generados, la confiabilidad de estos dos modelos predictivos es alta, aun cuando se tomó una

muestra de tiempo en la cual la fluctuación del indicador se pudo ver afectada por el impacto de la pandemia de COVID-19.

Figura 5. Resultados IBR



Fuente: elaboración propia.

Análisis de los títulos de deuda pública (TES 10 años)

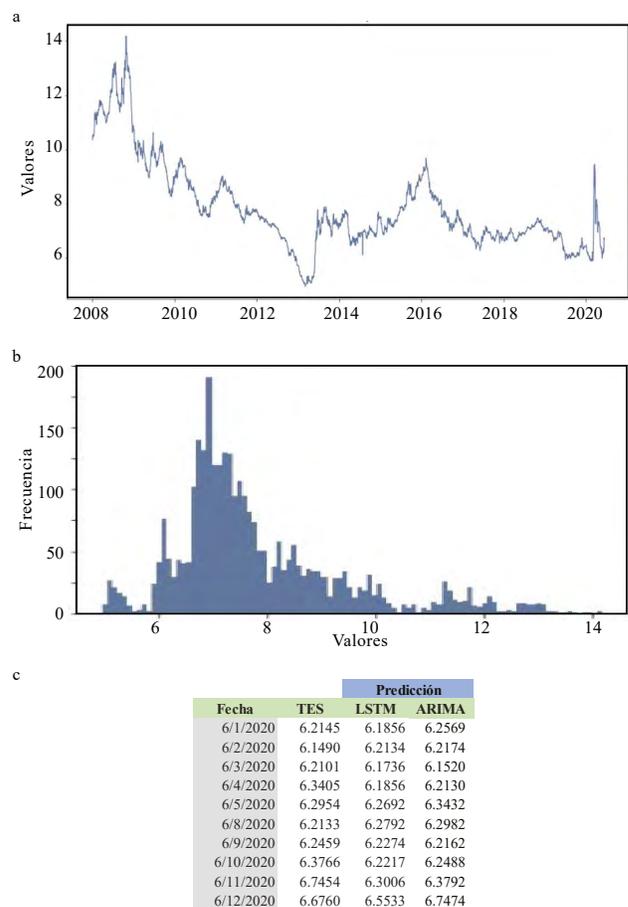
La importancia de comparar la tasa de rendimiento de los TES con el IBR radica en que los TES son el título de mayor inversión en el mercado de capitales, siendo los bancos quienes dan liquidez a los bonos a través del programa de creadores de mercado. Por tal razón, la emisión de bonos se relaciona en cierto modo con el IBR; además de estar estructurados sobre un cupón que corresponde a una tasa de interés.

Durante el período 2009-2015, el comportamiento de los TES presentó una caída en el rendimiento. En 2020, motivado por la pandemia, se evidencia un incremento en el rendimiento hasta el 24 de marzo, momento

en el cual nuevamente comenzó a descender, hasta alcanzar un mínimo de 5,985 % (figura 6a).

La frecuencia de los datos determina que la mayoría de los TES se encuentra en un rango entre 6-8 %, fluctuando entre 4,94651388919636 y 14,18142693689 % (figura 6b), con una media = 7,8261, mediana = 7,3593, desviación estándar = 1,6028 y varianza = 2,5691. Al igual que se hizo con los datos del IBR, se compararon los datos de los TES con los modelos predictivos Red LSTM y ARIMA, generando resultados que confirman su confiabilidad tras las predicciones con ambos modelos (figura 6c).

Figura 6. Resultados TES



Fuente: elaboración propia.

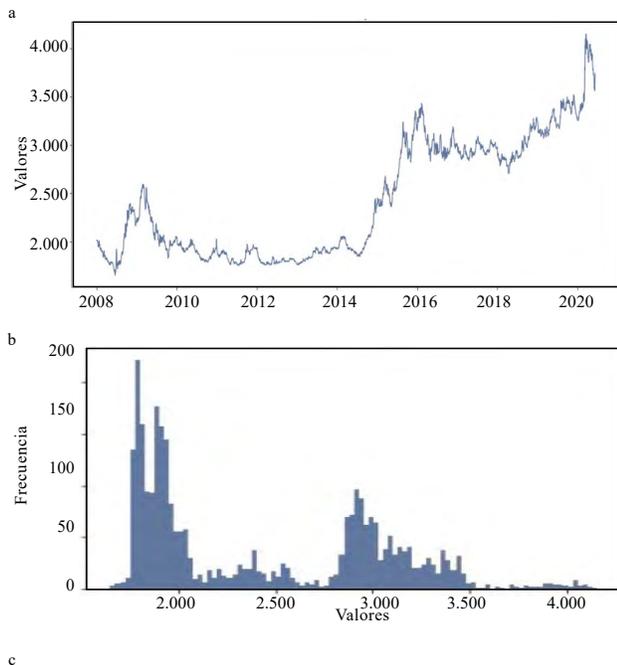
Análisis de la tasa representativa del mercado (TRM)

La TRM ha fluctuado por diferentes motivos, tanto globales como de la economía colombiana. El comportamiento de la TRM durante el período analizado

presenta una tendencia alcista, aunque cabe destacar que este crecimiento se presenta de manera más visible desde finales de 2015, como se muestra en la figura 7a. Además, es preciso mencionar que, por motivos de la pandemia, en el primer semestre de 2020 se llegó a un valor máximo de la TRM de 4.153,91 pesos colombianos (COP) por cada dólar estadounidense (USD).

Por otro lado, la frecuencia de los datos de la TRM determina que la mayoría de estos se encuentran en un rango entre COP 1.800 y 2.100, fluctuando entre COP 1.625,41 y 4.153,9, con una media = 2.423,7979 y una mediana = 2.158,5 (figura 7b). Finalmente, al igual que con los datos del IBR y los TES, se compararon los datos de la TRM con los modelos predictivos Red LSTM y ARIMA, generando una confiabilidad aceptable en ambos casos (figura 7c).

Figura 7. Resultados TRM



Fuente: elaboración propia.

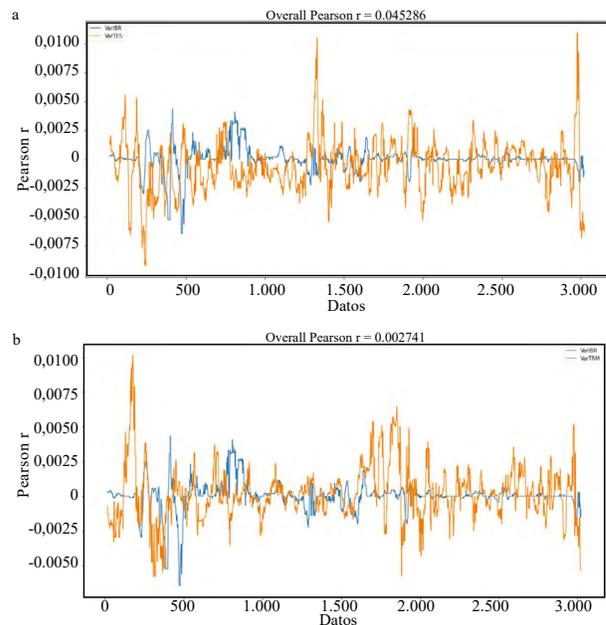
Análisis de correlación

Se determinó el coeficiente de correlación con el método Pearson y se analizaron tres variables cuantitativas (IBR, TES y TRM), las cuales fueron transformadas previamente a tasa de variación y permitieron obtener los siguientes resultados (figura 8):

- Análisis IBR/TES: $\rho = 0,045286$
- Análisis IBR/TRM: $\rho = 0,002741$

En los resultados se evidencia que no se cumple el supuesto de normalidad bivariada conjunta debido a que la variable ancho no se ajusta a la distribución gaussiana (o normal). Así, es posible que exista una relación no lineal entre las variables evaluadas debido a su cercanía a 0.

Figura 8. Coeficiente de correlación Pearson: a. Coeficiente para IBR/TES, b. Coeficiente para IBR/TRM



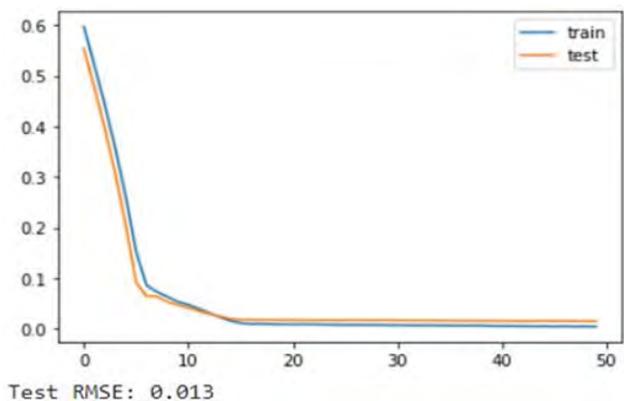
Fuente: elaboración propia.

Análisis del error del modelo de la red LSTM supervisada

Con el modelo ajustado y los pronósticos del conjunto de datos de prueba, se calculó la puntuación del error del modelo empleando el RMSE. El resultado obtenido fue de 0,013 (figura 9), lo que confirma la eficiencia de la red LSTM en la predicción. Para este caso, se utilizó

la versión eficiente de Adam del descenso de gradiente estocástico y la función de pérdida de MAE.

Figura 9. Gráfico de pérdida en el entrenamiento y conjunto de datos de prueba



Fuente: elaboración propia.

DISCUSIÓN Y CONCLUSIÓN

Teniendo en cuenta el objetivo principal de la investigación, se puede afirmar que la fiabilidad de los datos obtenidos con el modelo propuesto es aceptable (red LSTM). Los resultados contemplan un margen de error bajo entre los datos reales y predichos, observando una alta calidad en los datos generados. Desde el punto de vista del despliegue de la metodología CRISP-DM, se puede determinar que es factible llevar a cabo investigaciones tanto en ambientes financieros y económicos —en los que habitualmente se usan metodologías econométricas— como en otras áreas, permitiendo producir resultados de calidad. Si bien algunos pasos de la metodología son de alta complejidad y conllevan mucho tiempo, como la extracción y limpieza de datos o la elección del modelo, esta brinda confiabilidad en el resultado final.

Con el desarrollo paso a paso de la metodología CRISP-DM y su ciclo de vida para proyectos de minería de datos, se pudo evidenciar todo el procedimiento y el tratamiento de los datos, desde su comprensión hasta el despliegue, incluyendo fases vitales como el

modelado de datos, en el que se determinó la predicción del IBR con ayuda de una red LSTM multivariable supervisada, así como su RMSE.

Después de comprobar la eficiencia en la predicción de la temporalidad del indicador a través del método tradicionalmente empleado para la predicción de series temporales ARIMA y someter los datos a un sistema de red LSTM multivariable con tres entradas y una salida, determinando además su RMSE, concluimos que los resultados obtenidos con la red LSTM y su procesamiento interno de cuatro puertas (incluyendo una puerta de olvido) robustece la predicción, haciendo que esta sea altamente confiable. Estos resultados son semejantes a los datos obtenidos con otro modelo predictivo. Además, el error es bajo al comparar el valor observado con el conocido.

Al fijar los datos de entrada para la red LSTM multivariable, además de determinar el IBR, se buscó identificar la relación de este indicador con los TES y la TRM durante un período de 10 años con tasa de variación, donde igualmente fue posible determinar la eficiencia del modelo para una red LSTM simple, logrando obtener buenos resultados, en comparación con ARIMA. Finalmente se analizó la correlación de estas tres variables (IBR, TES y TRM) a través del método de Pearson, logrando evidenciar una correlación positiva muy baja, lo cual señala la posibilidad de que estas variables se relacionen entre sí de una forma no lineal.

Los resultados obtenidos con respecto al comportamiento de la predicción mediante el uso de la red LSTM y su sistema de puertas (que proporciona una fuente de memoria y de información procesada por unidades o celdas que se adaptan) son claros y abren una puerta para futuras investigaciones con datos financieros y fines predictivos. Al observar la baja correlación entre la IBR, los TES y la TRM, se podría considerar la realización de nuevos estudios con métodos de aprendizaje automático supervisado o no supervisado, con el fin de lograr una mejor comprensión del comportamiento de estas variables y de los efectos de su posible estacionalidad.

REFERENCIAS

- Abrantes-Metz, R. M., Villas-Boas, S. B., & Judge, G. (2011). Tracking the LIBOR rate. *Applied Economics Letters*, 18(10), 893-899. <https://doi.org/10.1080/13504851.2010.515197>
- Bükk, A., & Johansson, R. (2020). *Vehicle trajectory prediction using recurrent LSTM neural networks* (tesis de maestría, Chalmers University of Technology). https://odr.chalmers.se/bitstream/20.500.12380/300726/1/B%C3%BCKk_Johansson_2020.pdf
- Cardozo, P., & Murcia, A. (2014). *El IBR y la DTF como indicadores de referencia*. Reportes del Emisor n.º 187. Banco de la República de Colombia.
- Castillo-Mur, A. (2020). *Las criptomonedas en el sistema financiero internacional. Un análisis predictivo mediante técnicas de machine learning (LSTM)* (tesis de pregrado Universidad de Barcelona/Universidad Politécnica de Catalunya). Dipòsit Digital de la Universitat de Barcelona. <http://diposit.ub.edu/dspace/handle/2445/171827>
- Dmitrieva, L., Kuperin, Y., & Soroka, I. (2002). Neural network prediction of short-term dynamics of futures on Deutsche mark, LIBOR and S&P500. En P. M. A. Sloot, A. G. Hoekstra, C. J. K. Tan, & J. J. Dongarra (eds.), *Computational Science — iccs 2002. Lecture Notes in Computer Science* (pp. 1201-1208). Springer. https://doi.org/10.1007/3-540-47789-6_127
- Duffie, D., & Stein, J. C. (2015). Reforming LIBOR and other financial market benchmarks. *Journal of Economic Perspectives*, 29(2), 191-212. <https://doi.org/10.1257/jep.29.2.191>
- Fagerström, J., Bång, M., Wilhelms, D., & Chew, M. S. (2019). LiSep LSTM: A machine learning algorithm for early detection of septic shock. *Scientific Reports*, 9(1), 15132. <https://doi.org/10.1038/s41598-019-51219-4>
- Jamshidian, F. (1997). LIBOR and swap market models and measures. *Finance and Stochastics*, 1(4), 293-330. <https://doi.org/10.1007/s007800050026>
- Kumar, J., Goomer, R., & Singh, A. K. (2018). Long short-term memory recurrent neural network (LSTM-RNN) based workload forecasting model for cloud datacenters. *Procedia Computer Science*, 125, 676-682. <https://doi.org/10.1016/j.procs.2017.12.087>
- Leal, F., Molina, C., & Zilberman, E. (2020). *Proyección de la inflación en Chile con métodos de machine learning*. Banco Central de Chile
- Li, D., Zhang, Y., Gong, D., Pan, L., & Zhao, Y. (2020). Gas data prediction based on LSTM neural network. *MS&E*, 750(1), 012175. <https://doi.org/10.1088/1757-899X/750/1/012175>
- Liu, P., Qiu, X., & Huang, X. (2016). Recurrent neural network for text classification with multi-task learning. *arXiv preprint arXiv*, 1605.05101.
- Pai, P. F., & Lin, C. S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*, 33(6), 497-505. <https://doi.org/10.1016/j.omega.2004.07.024>
- Pastor-Cobo, M., & Torre-Barrio, P. *Elaboración de un algoritmo predictivo sobre la evolución del precio de las criptomonedas* (tesis de pregrado, Universidad Complutense de Madrid). Repositorio Universidad Complutense de Madrid.
- Qiu, J., Wang, B., & Zhou, C. (2020). Forecasting stock prices with long-short term memory neural network based on attention mechanism. *PLoS One*, 15(1), e0227222. <https://doi.org/10.1371/journal.pone.0227222>
- Restrepo, L. F., & González, J. (2007). From Pearson to Spearman. *Revista Colombiana de Ciencias Pecuarias*, 20(2), 183-192.
- Robles, H. P. (2019). Uso de redes neuronales para el pronóstico de la inflación en México. *Almanaque Económico*, 1, 69-74.
- Rodríguez-Vargas, A. (2020). *Forecasting Costa Rican inflation with machine learning methods*. Banco Central de Costa Rica.

- Tan, X. (2019). LIBOR prediction using genetic algorithm and genetic algorithm integrated with recurrent neural network. *2019 Global Conference for Advancement in Technology (GCAT)*. Bangalore, India. <https://doi.org/10.1109/GCAT47503.2019.8978299>
- Villada, F., Muñoz, N., & García, E. (2012). Aplicación de las redes neuronales al pronóstico de precios en el mercado de valores. *Información Tecnológica*, 23(4), 11-20. <https://doi.org/10.4067/S0718-07642012000400003>
- Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining* (pp. 29-39).
- Zhang, X., Liang, X., Zhiyuli, A., Zhang, S., Xu, R., & Wu, B. (2019). AT-LSTM: An attention-based LSTM model for financial time series prediction. *IOP Conference Series: Materials science and engineering*, 569, 052037. <https://doi.org/10.1088/1757-899X/569/5/052037>

Analítica de grafos para identificar entidades relevantes y comunidades en Mercado Libre: un estudio de caso

Use of Graph Analytics to Identify Relevant Entities and Communities in Mercado Libre: A Case Study

Gerson Enrique Carrillo Gelvez^{ac}, Ixent Galpin^{bd}

^a Facultad de Ciencias Naturales e Ingeniería, Universidad de Bogotá Jorge Tadeo Lozano, Colombia

^b Grupo de Investigación Ingeniería de Datos y Sistemas Inteligentes (ID&SI), Facultad de Ciencias Naturales e Ingeniería, Universidad de Bogotá Jorge Tadeo Lozano, Colombia

^c gersonc.carrillog@utadeo.edu.co | <https://orcid.org/0000-0001-6959-0318>

^d <https://orcid.org/0000-0001-7020-6328>

RESUMEN

Este artículo representa la información disponible en bases de datos no relacionales, aprovechando los beneficios de escalabilidad, alta disponibilidad, resiliencia y facilidad proporcionados por estas. Así mismo, se da a conocer una serie de algoritmos suministrados por el motor de bases de datos de grafos Neo4j para computar métricas de grafos, nodos y relaciones. En primer lugar, se consolida un conjunto de datos públicos tomado del sistema de ventas online de Mercado Libre. Posteriormente, se modelan los datos obtenidos en un esquema de grafos que tiene como nodos a los usuarios, quienes pueden ser vendedores, compradores, productos y sus características. Como siguiente paso, se aplican algoritmos que calculan métricas del grafo, junto con sus nodos y relaciones, visualizando de esta manera los resultados obtenidos. Para finalizar, se identifican las categorías ofertadas más importantes, las comunidades existentes y los usuarios más influyentes.

Palabras clave: base de datos de grafos, analítica de grafos, NOSQL, métricas de centralidad, detección de comunidades.

Citation: Carrillo-Gelvez, G. E. y Galpin, I. (2021). Analítica de grafos para identificar entidades relevantes y comunidades en Mercado Libre: un estudio de caso. *Mutis*, 11(1), 77-95.
<https://doi.org/10.21789/22561498.1740>

Recibido: Enero 18, 2021. **Aceptado:** Marzo 30, 2021.

Copyright: ©2021. Chávez-Plazas, Y. A., Ramírez-Mahecha, M. C. y Camacho-Kurmen, J. E. (2021). This is an open-access article, which permits unrestricted use, distributions and reproduction in any medium, provided the original author and source are credited.

Competing Interests: The authors have no conflict of interest.

ABSTRACT

This article represents the information available in non-relational databases, taking advantage of their scalability, high availability, resilience, and ease of development. This work also describes some algorithms provided by the Neo4j graph database engine to compute graph, node and relationship metrics. To do this, we first consolidate a data set obtained from Mercado Libre online sales system. Subsequently, the data is cast into a graph schema that considers users



as nodes. Such users can be sellers or buyers, products and their characteristics. Afterward, we applied the algorithms that calculate metrics from the graph, as well as its nodes and relationships, thus displaying the results obtained. Finally, we identify the most important categories offered, along with the most influential communities and users.

Keywords: Graph database, graph analytics, NoSQL, centrality metrics, community detection.

INTRODUCCIÓN

Desde la década de 1970, las bases de datos relacionales han funcionado de manera exitosa con datos que encajan en el formato de tablas, columnas y filas. No obstante, las consultas sobre conjuntos de datos con entidades altamente interrelacionadas son complejas de expresar en SQL, puesto que a menudo requieren de un alto número de *joins* y tienden a consumir excesivos recursos de ejecución. De ahí el surgimiento de las bases de datos de grafos, las cuales están enfocadas hacia datos conectados y permiten a las organizaciones comprender la gran cantidad de conexiones existentes entre entidades (Svensson, 2020). El uso de las bases de datos NoSQL en proyectos de desarrollo de software ha ido incrementando (Pragma, 2018). Por este motivo, es importante explorar ejemplos de cómo se puede modelar una base de datos de grafos mediante la aplicación de algoritmos, teniendo en cuenta que los volúmenes de datos de los sistemas crecen cada día más y más.

DB-Engines (2020a) clasifica los sistemas de administración de bases de datos según su popularidad en un ranking que se actualiza mensualmente. Para el mes de octubre de 2020, las bases de datos relacionales ocuparon los 4 primeros puestos (Oracle, MySQL, Microsoft SQL Server y PostgreSQL), mientras que el quinto lugar fue ocupado por la base de datos de documentos MongoDB. El puesto número 21 del ranking fue ocupado por la base de datos orientada a grafos Neo4j, lo que indica que este tipo de bases viene ganando popularidad.

Cuando se crea un nuevo sistema de información se seleccionan motores de bases de datos diferentes a los orientados a grafos. Posiblemente, esto se debe a que se carece de conocimiento sobre las propiedades y los beneficios que este tipo de motores pueden traer a la solución. Este mismo desconocimiento imposibilita identificar que se puede contar

con conjuntos de datos limitados para evaluar y probar las características de las bases de datos de grafos, dentro de las que Neo4j, Microsoft Azure Cosmos DB, ArangoDB, OrientDB y Virtuoso son las 5 más populares (DB-Engines, 2020b). A pesar de esto, los grafos han sido usados para analizar diversos temas, desde fraudes financieros hasta el rendimiento deportivo, lo que parece indicar que estas bases de datos tienen un alto potencial que en la actualidad no está siendo explotado.

El trabajo descrito en este artículo consolida información de Mercado Libre, una plataforma de comercio electrónico de ventas en línea, con información de vendedores, compradores, productos y las características de estos. Mercado Libre es la compañía argentina de comercio electrónico que se convirtió en la empresa latinoamericana más valiosa en términos de capitalización de mercado (Dinero, 2020). Debido a que no existe actualmente un conjunto de datos de una tienda en línea de productos de tecnología, la presente investigación puede ser de ayuda para una mejor comprensión sobre dichos datos, dado que el funcionamiento de las tiendas en línea es conocido por la gran mayoría de las personas y hace que los datos sean fácilmente entendibles.

Los componentes claves de esta investigación están orientados a entender cómo extraer información de una página web y cómo modelar dicha información en una base de datos de grafos, en este caso, Neo4j, con la cual se pueda hacer una representación abstracta que describe la organización de los sistemas de transporte, las interacciones humanas, las telecomunicaciones en redes y las tiendas en línea. A partir de lo anterior, es posible modelar un gran número de estructuras diferentes usando tan solo un paradigma, otorgando así un gran poder a la comunidad de analítica de datos. Para ello, se aplican algoritmos que proporcionan métricas sobre el grafo, los nodos y las relaciones. Estos algoritmos ayudan a detectar los nodos y las comunidades más influyentes en la información.

ESTADO DEL ARTE

Los algoritmos de detección de comunidades se enfocan principalmente en la agrupación de grafos (Scott, 2011) y la influencia de los usuarios (Scott & Carrington, 2011). Rossi y Ahmed (2015) crearon el primer repositorio de datos interactivo con un sitio web para descargar información o aplicar analítica de grafos sobre conjuntos de datos de diferentes categorías. Adicional a esto, los usuarios pueden discutir cada grafo, así como publicar observaciones y visualizaciones que facilitan la investigación científica.

El trabajo de Vicknair *et al.* (2010) realiza un comparativo entre una base de datos relacional y una de grafos. En general, la base de datos de grafos funciona mejor en las consultas de tipo estructural que la base de datos relacional. Así mismo, en búsquedas de caracteres de texto completo, la base de datos de grafos es significativamente mejor que la base de datos relacional. Esto se debe a que el mecanismo de indexación utilizado en la base de datos de grafos se basa en cadenas, lo que hace que las consultas numéricas sean menos eficientes. No obstante, es pertinente mencionar que un factor clave en la elección de un sistema de base de datos es la seguridad, frente a lo que la falta de soporte ofrecida por Neo4j se convierte en una limitación (Vicknair *et al.*, 2010).

En las comunidades financieras los fraudes de tipo bancario, seguros y comercio electrónico tienen comportamientos diferentes a los patrones de comportamiento “normal” (Das & Sisk, 2005). Aun así, estos sectores tienen en común que los fraudes se realizan de manera indirecta. Teniendo en cuenta que los estafadores son cada vez más sofisticados al momento de eludir los métodos tradicionales de detección de fraude, trabajando juntos y creando nuevas identidades, los grafos están diseñados para ver las relaciones de los usuarios de forma directa o indirecta, con lo que se ofrece una oportunidad significativa de aumentar los métodos existentes de detección de fraude (Sadowski & Rathle, 2014). Así, se da un nuevo enfoque a los métodos usados para identificar transacciones bancarias irregulares a partir del uso de propiedades de algoritmos de detección de comunidades. Al respecto, de acuerdo con los resultados del estudio de Molloy *et al.* (2017), es posible reducir sustancialmente los falsos positivos en la puntuación de fraude tradicional. Así mismo, Eboli (2007) presenta un enfoque

novedoso para el análisis de impagos de sistemas financieros, representando la información en un grafo con las pérdidas que se propagan en dicha red y calculando las pérdidas e incumplimientos de los usuarios.

En el análisis de las redes sociales el descubrimiento de comunidades es esencial. Por ello, se implementan algoritmos para el descubrimiento de comunidades en Neo4j, evaluando los resultados obtenidos a través de la comparación de métricas de similitud (Kanavos *et al.*, 2017). En este contexto, la descomposición de forma recursiva de un grafo permite el análisis de una red social, evidenciando que el algoritmo de detección de comunidades Louvain crea comunidades mejor distribuidas (Neo4j, 2020a). Por otra parte, se usan algoritmos de centralidad que permiten identificar los nodos más importantes de un grafo. Un ejemplo significativo es identificar la importancia de las páginas web y sus enlaces usando el algoritmo PageRank (Page *et al.*, 1999) para identificar usuarios influyentes en Twitter (Weng *et al.*, 2010) y sus variantes (Kleinberg, 1999).

Otro ejemplo es la popular serie *Game of Thrones*, que cuenta con una gran audiencia, la cual ha sido analizada mediante el esquema de grafos para evidenciar la importancia de los personajes según su participación en cada temporada (Neo4j, 2020b).

De otro lado, la teoría de grafos también se puede implementar en el análisis de rendimiento deportivo con el objetivo de evaluar el rendimiento individual y colectivo de los deportistas. Además, por medio de esta teoría se describen posibles limitaciones de estudios sobre el tema en redes sociales, proporcionando sugerencias para futuras investigaciones en el campo (Ribeiro *et al.*, 2017).

En el área de la salud, la conmoción cerebral relacionada con el deporte es un problema importante de salud pública. Sin embargo, se sabe poco sobre los cambios en redes cerebrales funcionales en adolescentes después de una lesión. Frente a este asunto, Virji-Babul *et al.* (2014) utilizan la teoría de grafos para evaluar los cambios en las propiedades de la red cerebral después de una conmoción cerebral en atletas adolescentes. Por su parte, Branting *et al.* (2016) han empleado algoritmos de grafos para identificar riesgo de fraude sanitario, calculando la similitud de actividades de atención médica, procedimientos y prescripción de medicamentos.

CONJUNTO DE DATOS DE MERCADO LIBRE

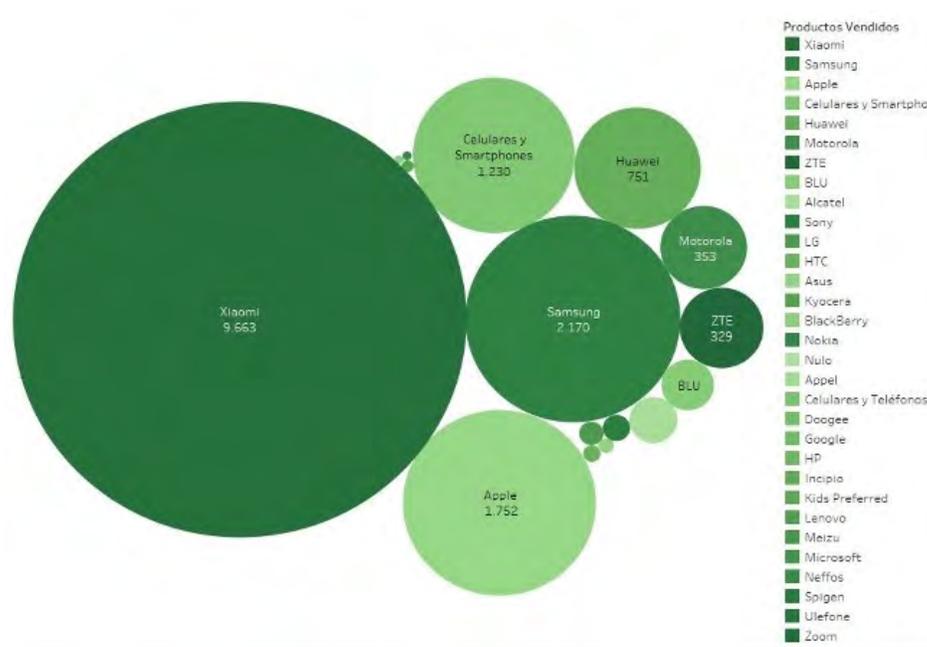
Los datos del presente estudio fueron extraídos de la página web Mercado Libre Colombia (<https://www.mercadolibre.com.co/>), actualizada a septiembre de 2020, y corresponden a información de productos, usuarios vendedores, usuarios compradores y opiniones que hacen parte de la categoría “tecnología” en la ciudad de Bogotá.

La información fue obtenida mediante técnicas de *webscraping* con el lenguaje de programación Python, usando la librería Selenium.¹ Con el objetivo de asegurar que la información tuviera un alcance definido, esta fue filtrada a partir del siguiente criterio: información de Bogotá (Colombia) dentro de la categoría “Tecnología, Celulares y Teléfonos, Celulares y Smartphones”. De esta manera, se obtuvieron 5.543 registros de productos, 1.096 registros de usuarios vendedores, 15.526 registros de usuarios compradores y 15.997 opiniones.

DESCRIPCIÓN DE LOS DATOS

Inicialmente, se realizó una exploración del conjunto de datos con el objetivo de entender el comportamiento de los productos vendidos y los productos ofertados. En la figura 1 se identifica que las categorías generales más vendidas (a la fecha de consulta) son Xiaomi (9.663 productos) y Samsung (2.170 productos). Comprobando el detalle, en la figura 2 se observa que los productos de categoría Note 8 y Note 8 Pro ocupan los dos primeros lugares.

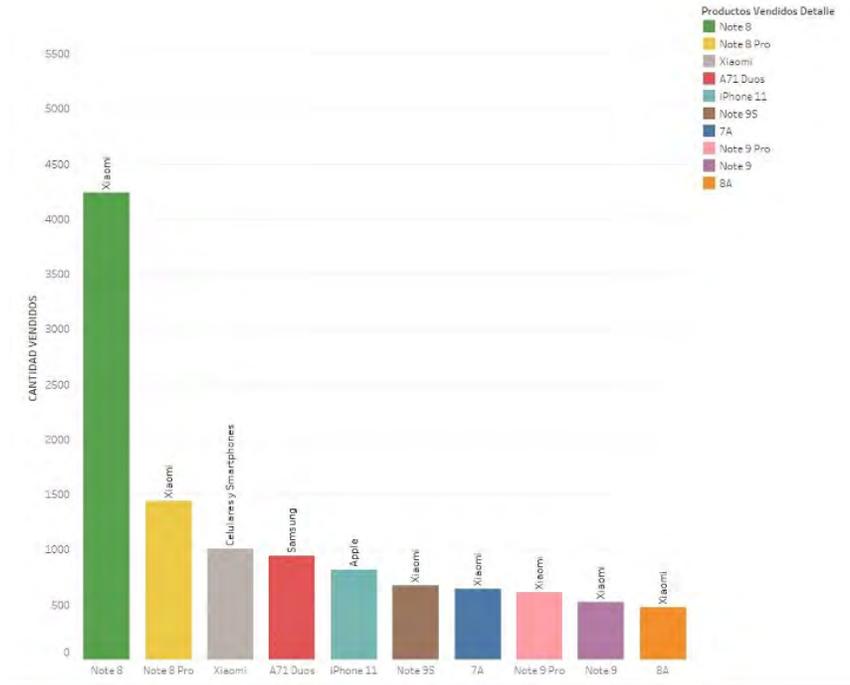
Figura 1. Descripción de los datos: productos vendidos



Fuente: elaboración propia.

1 <https://pypi.org/project/selenium/>

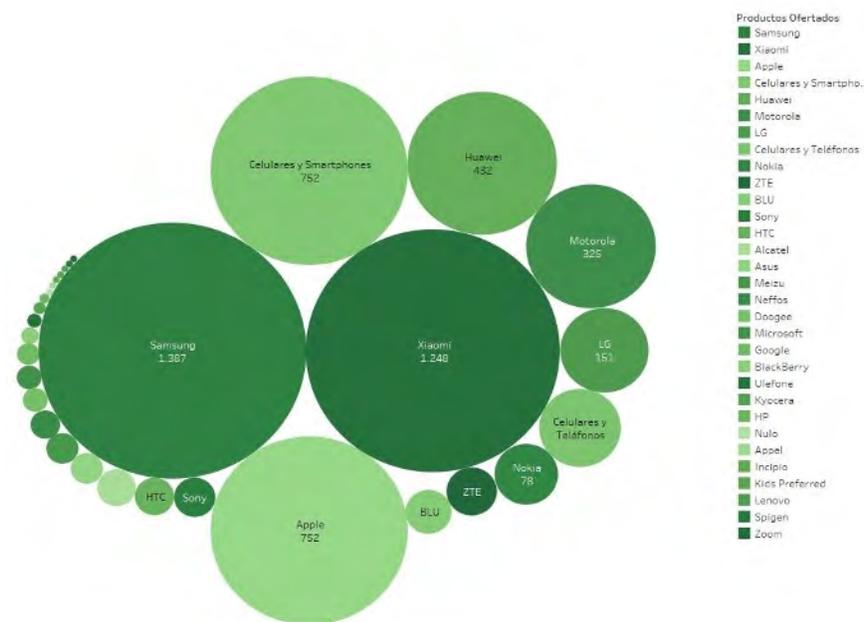
Figura 2. Descripción de los datos: detalle de los productos vendidos



Fuente: elaboración propia.

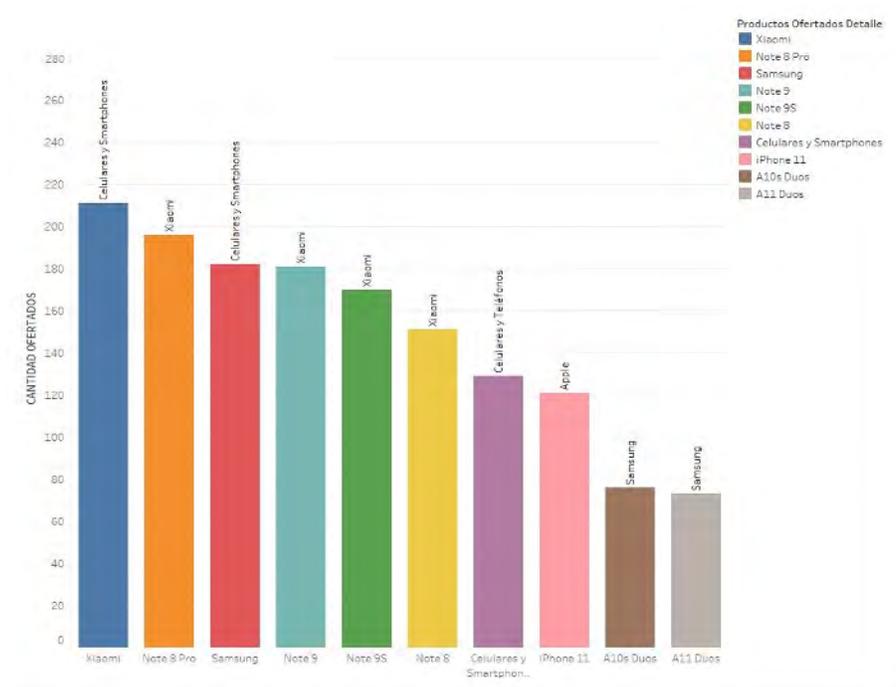
En la figura 3 se observa que las categorías generales más ofertadas también corresponden a Samsung (1.387 productos) y Xiaomi (1.248 productos). Viendo el detalle de estas categorías, en la figura 4 se aprecia que Xiaomi y Note 8 Pro siguen liderando los dos primeros lugares.

Figura 3. Descripción de los datos: productos ofertados



Fuente: elaboración propia.

Figura 4. Descripción de los datos: detalle de los productos ofertados



Fuente: elaboración propia.

La figura 5 evidencia que la cantidad de productos vendidos está relacionada con la antigüedad del vendedor en la plataforma de Mercado Libre, como era de esperarse. Sin embargo, una notable excepción es el vendedor “Celulares 99”, quien inició actividades el 1 de agosto de 2020 y se ubica en el puesto número cinco de este ranking.

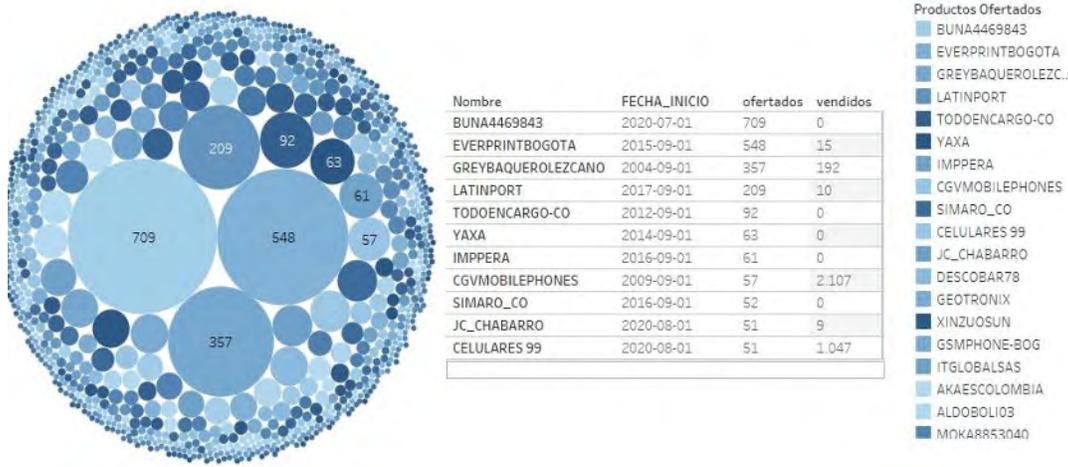
Figura 5. Descripción de los datos: cantidad de productos vendidos



Fuente: elaboración propia.

De otro lado, en la figura 6 se observa que la cantidad de productos ofertados no está relacionada con la cantidad de productos vendidos ni con la fecha de inicio del vendedor en Mercado Libre, dado que el vendedor con más productos ofertados inició actividades el 1 de julio de 2020.

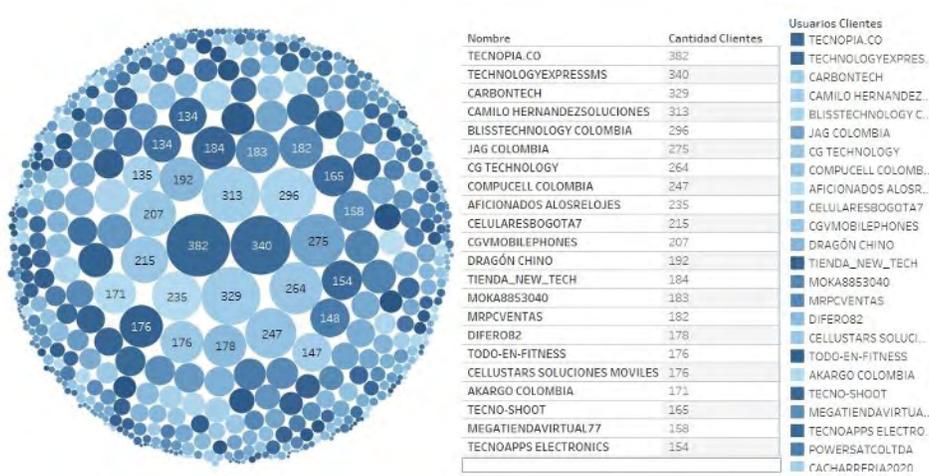
Figura 6. Descripción de los datos: cantidad de productos ofertados



Fuente: elaboración propia.

La figura 7 muestra que existe un conjunto de vendedores constantes en Mercado Libre, lo cual se puede deducir por la cantidad de clientes que estos tienen. En contraste, un mayor grupo de vendedores cuenta con pocos clientes, indicando que se trata de vendedores esporádicos en esta plataforma.

Figura 7. Descripción de los datos: usuarios vendedores



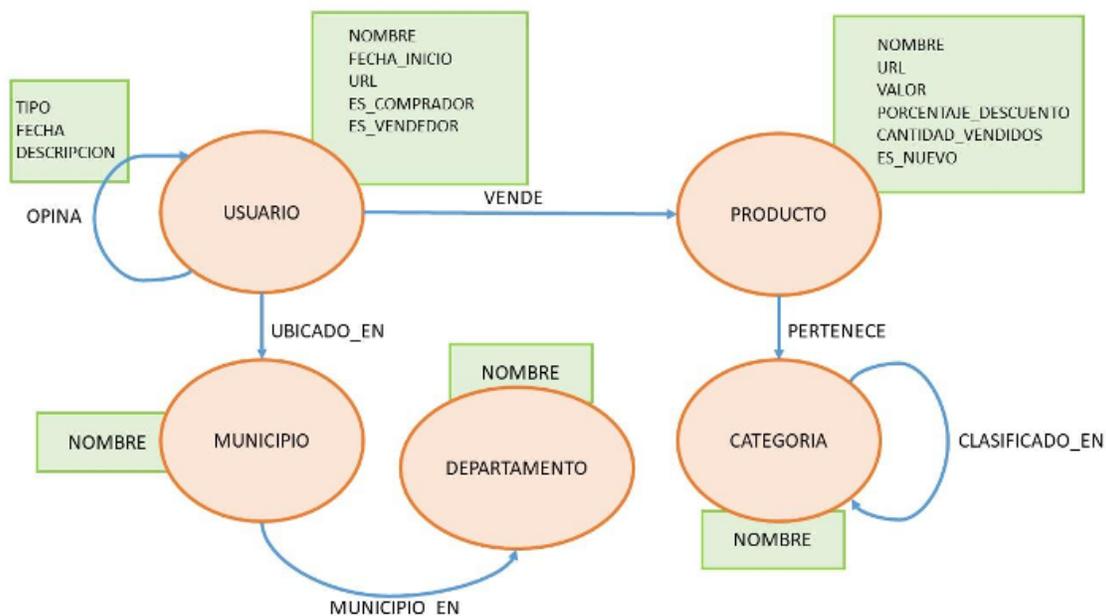
Fuente: elaboración propia.

ESQUEMA DEL GRAFO

De acuerdo con el modelo de datos de Neo4j, un grafo es una composición de un conjunto de objetos, conocidos como nodos, que se relacionan con otros (nodos) a través de un conjunto de conexiones, conocidas como aristas (Neo4j, 2020c). Los nodos se utilizan para representar entidades, mientras que las etiquetas sirven para agrupar nodos en conjuntos donde todos los nodos que tienen determinada etiqueta pertenecen al mismo conjunto. Así, un nodo puede estar asociado a cero, una o varias etiquetas. De otro lado, una relación expresa conexión entre dos nodos. En Neo4j, cada relación tiene exactamente un tipo. Las relaciones, a su vez, organizan a los nodos en estructuras, lo que permite que un grafo represente una lista, un árbol, un mapa o una entidad compuesta (Neo4j, 2020c).

En la figura 8 se definen nodos y relaciones que comprenden el modelo de datos del conjunto de datos de Mercado Libre, mientras que la tabla 1 presenta el respectivo diccionario de datos. Los nodos con etiqueta DEPARTAMENTO representan los departamentos registrados en el sistema. Los nodos con etiqueta MUNICIPIO se conectan con el departamento mediante la relación MUNICIPIO_EN. Los nodos con etiqueta USUARIO contiene los datos de vendedores y compradores, donde un usuario puede tener uno o ambos roles. Los usuarios están situados en un municipio (a través de la relación UBICADO_EN), opinan sobre otros usuarios (a través de la relación OPINA) y venden productos (a través de la relación VENDE). Los nodos con etiqueta PRODUCTO contienen los datos de los productos ofertados por los vendedores y están clasificados dentro de una categoría mediante la relación PERTENECE. Por último, los nodos con etiqueta CATEGORIA clasifican a los productos, considerando que una categoría puede estar clasificada dentro de otra, generando una relación jerárquica (relación CLASIFICADO_EN).

Figura 8. Esquema de grafos



Fuente: elaboración propia.

Tabla 1. Diccionario de datos

Tipo objeto	Nombre objeto	Descripción objeto	Nombre campo	Descripción campo	Tipo dato
Nodo	DEPARTAMENTO	Departamento donde los usuarios están ubicados	CÓDIGO	Código único del departamento	STRING
			NOMBRE	Nombre del departamento	STRING
Nodo	MUNICIPIO	Municipio donde los usuarios están ubicados	CÓDIGO	Código único del municipio	STRING
			NOMBRE	Nombre del municipio	STRING
Nodo	USUARIO	Usuarios vendedores o compradores	NOMBRE	Nombre del usuario	STRING
			FECHA_INICIO	Fecha de inicio del usuario en el sistema	DATE
			URL	URL del perfil del usuario	STRING
			ES_COMPRADOR	Identifica si el usuario es comprador	BOOLEAN
			ES_VENDEDOR	Identifica si el usuario es vendedor	BOOLEAN
Nodo	PRODUCTO	Productos ofertados	NOMBRE	Nombre del producto	STRING
			URL	URL del producto ofertado	STRING
			VALOR	Valor del producto	STRING
			PORCENTAJE_DESCUENTO	Porcentaje de descuento del producto	STRING
			CANTIDAD_VENDIDO	Cantid de productos vendidos	STRING
			ES_NUEVO	Identifica si el producto es nuevo	BOOLEAN
Nodo	CATEGORÍA	Clasificación de los productos	CÓDIGO	Código único del departamento	STRING
			NOMBRE	Nombre del departamento	STRING
Relación	MUNICIPIO_EN	Relación entre los nodos MUNICIPIO y DEPARTAMENTO			
Relación	UBICADO_EN	Relación del municipio donde se ubica un usuario			
Relación	OPINA	Opinión que da un usuario comprador a un usuario vendedor	TIPO	Tipo de opinión	STRING
			FECHA	Fecha de opinión	DAT
			DESCRIPCIÓN	Descipción de opinión	STRING
Relación	VENDE	Relación entre el producto ofertado y el usuario vendedor			
Relación	PERTENECE	Relación que clasifica a un producto			
Relación	CLASIFICADO_EN	Relación jerárquica de las categorías			

Fuente: elaboración propia.

IMPORTACIÓN DE LOS DATOS EN NEO4J

La información recolectada con técnicas de *webscraping* fue guardada en forma tabular usando archivos con formato csv. Neo4j, con su lenguaje Cypher, proporciona una opción para facilitar la importación de los datos contenidos en los archivos a una estructura de grafos.

Es importante desarrollar los siguientes pasos para cargar los datos a Neo4j:

- Contar con un motor de bases de datos Neo4j instalado y en correcto funcionamiento.

- Los datos se pueden descargar en la url https://github.com/gersongelvez/TESIS_MAESTRIA/tree/master/DATOS (Carrillo-Gelvez, 2019).
- Usar la herramienta de importación de archivos con formato CSV.
- Ejecutar los comandos de importación uno a uno en la consola de Neo4j.

APLICACIÓN DE CONSULTAS EN LENGUAJE CYPHER

Neo4j permite la realización de consultas usando Cypher, un lenguaje declarativo inspirado en SPARQL (Pérez *et al.*, 2009) que permite expresar la coincidencia de patrones en grafos. En esta sección se presentan algunas consultas sobre el esquema del grafo definido para Mercado Libre con el fin de ilustrar las funcionalidades de Cypher.

Listado de departamentos y municipios

La consulta en lenguaje Cypher para listar los departamentos y municipios es:

```
MATCH (M:MUNICIPIO)-[:MUNICIPIO_EN]->(D:DEPARTAMENTO)
RETURN D.NOMBRE AS DEPARTAMENTO, M.NOMBRE AS MUNICIPIO
```

Esta consulta busca el patrón en el grafo que consisten en un nodo M con la etiqueta MUNICIPIO y un nodo D con la etiqueta DEPARTAMENTO, que se conectan a través de una relación de tipo MUNICIPIO_EN.

La tabla 2 muestra que existen departamentos y municipios diferentes a Bogotá, lo que quiere decir que a pesar de que solo se recolectó información de productos de tecnología en Bogotá hay vendedores de diferentes partes de Colombia dentro de esta categoría. Además, la página de Mercado Libre registra a Bogotá como un departamento y a sus localidades como municipios, por lo que resulta necesario aplicar un proceso de calidad de datos para estandarizar la información.

Tabla 2. Departamentos y municipios

Departamento	Municipio
ATLÁNTICO	BARRANQUILLA
BOGOTÁ DC	USME
BOGOTÁ DC	USAQUÉN

Departamento	Municipio
BOGOTÁ DC	TUNJUELITO
BOGOTÁ DC	TEUSAQUILLO
BOGOTÁ DC	SUMAPAZ
BOGOTÁ DC	SUBA
BOGOTÁ DC	SANTA FE
BOGOTÁ DC	SAN CRISTOBAL SUR
BOGOTÁ DC	RAFAEL URIBE URIBE
BOGOTÁ DC	PUENTE ARANDA
BOGOTÁ DC	MÁRTIRES
BOGOTÁ DC	LA CANDELARIA
BOGOTÁ DC	KENNEDY
BOGOTÁ DC	FONTIBÓN
BOGOTÁ DC	ENGATIVÁ
BOGOTÁ DC	CIUDAD BOLIVAR
BOGOTÁ DC	CHAPINERO
BOGOTÁ DC	BOSA
BOGOTÁ DC	BARRIOS UNIDOS
BOGOTÁ DC	ANTONIO NARIÑO
CESAR	AGUACHICA
CUNDINAMARCA	SOACHA
CUNDINAMARCA	MOSQUERA
CUNDINAMARCA	FUNZA

Fuente: elaboración propia.

Cantidad de usuarios por departamento y municipio

La consulta en lenguaje Cypher para listar los 10 municipios y su respectivo departamento con más usuarios vendedores es:

```
MATCH (U:USUARIO)-[:UBICADO_EN]->(M:MUNICIPIO)-[:MUNICIPIO_EN]->(D:DEPARTAMENTO)
RETURN D.NOMBRE AS DEPARTAMENTO, M.NOMBRE AS MUNICIPIO,
COUNT(U) AS CANTIDAD ORDER BY CANTIDAD DESC LIMIT 10
```

Esta consulta muestra el uso de la función de agregación COUNT, similar a la del lenguaje SQL, que en este caso realiza un conteo de los usuarios U y les asigna el alias CANTIDAD. El resultado se ordena de forma descendente y se limita a diez filas.

La tabla 3 muestra que el municipio de Suba contiene la mayor cantidad de usuarios vendedores. Por otro lado, se identifica que los nombres de los municipios necesitan un proceso de calidad de datos debido a algunos errores de ortografía en sus nombres.

Tabla 3. Usuarios por departamento y municipio

Departamento	Municipio	Cantidad
BOGOTÁ DC	SUBA	130
BOGOTÁ DC	KENNEDY	115
BOGOTÁ DC	ENGATIVÁ	96
BOGOTÁ DC	CHAPINERO	86
BOGOTÁ DC	USAQUÉN	85
BOGOTÁ DC	PUENTE ARANDA	58
BOGOTÁ DC	BOSA	48
BOGOTÁ DC	FONTIBÓN	47
BOGOTÁ DC	TEUSAQUILLO	46

Fuente: elaboración propia.

Tabla 4. Productos de Apple más vendidos

Vendedor	Producto	Categoría	Valor (USD)	Cantidad vendidos
Celulares 99	Celular iPhone 11 128gb nuevo 100 % original y sellada	iPhone 11	943,81	203
Celulares 99	Celular iPhone 11 64gb sellado nuevo 4g	iPhone 11	850,77	193
Mobile Connection	iPhone 7 32gb 4g LTE 12mp 4k 3d touch Garantía + obsequios	iPhone 7	306,46	114
Celulares 99	Celular iPhone SE 2020 64gb chip A13	SE	533,58	100
Gmsphone-bog	iPhone 11 64gb sellado entrega inmediata 1 año garantía	iPhone 11	848,28	80

Fuente: elaboración propia.

Categorías con el máximo número de niveles

La consulta en lenguaje Cypher para mostrar las cinco categorías con el máximo número de niveles es:

```
MATCH path = (HIJA:CATEGORIA)-[:CLASIFICADO_EN *1..]->(PADRE:CATEGORIA)
RETURN PADRE.NOMBRE AS CATEGORIA, MAX(length(path)) as depth
ORDER BY depth DESC, CATEGORIA LIMIT 5
```

Esta consulta busca el patrón en el grafo donde aparece un camino o una (o más) arista del tipo CLASIFICADO_EN y lo asigna a la variable *path*. Posteriormente, la función *length* permite contar la longitud del camino en el grafo. Así, la tabla 5 indica que el máximo número de niveles que tiene la jerarquía de categorías es cuatro.

Tabla 5. Categorías con el máximo número de niveles

Categoría	Niveles
Calulares y teléfonos	4
Celulares y smartphones	3
Alcatel	2
Apple	2
Apple	2

Fuente: elaboración propia.

Productos de Apple más vendidos

La consulta en lenguaje Cypher para listar los 5 productos de Apple más vendidos y sus características es:

```
MATCH (U:USUARIO)-[:VENDE]->(P:PRODUCTO)-[:PERTENECE]->(C:CATEGORIA)
WHERE C.CODIGO CONTAINS 'Apple'
RETURN U.NOMBRE AS VENDEDOR, P.NOMBRE AS PRODUCTO, C.NOMBRE AS CATEGORIA,
toInteger(P.VALOR)/3703.7 AS "VALOR (USD)", toInteger(P.CANTIDAD_VENDIDOS)
AS CANTIDAD_VENDIDOS ORDER BY CANTIDAD_VENDIDOS DESC LIMIT 5
```

En esta consulta la cláusula WHERE permite filtrar por atributos de los nodos, en este caso estipulado que solo se deben incluir usuarios que venden productos de la categoría Apple. La tabla 4 indica que el vendedor "Celulares 99" es quien ha vendido más productos de la categoría Apple.

APLICACIÓN DE ANALÍTICA DE GRAFOS

La librería Graph Data Science (GDS)² de Neo4j ofrece una serie de algoritmos para aplicar analítica de grafos sobre el conjunto de datos de Mercado Libre. En esta sección se describen los conceptos que se deben tener en cuenta para ello.

Algoritmos

Los algoritmos se utilizan para calcular métricas de grafos, nodos o relaciones. Estos pueden proporcionar información sobre entidades relevantes en el grafo (centralidad y clasificación) o estructuras inherentes, como la detección de comunidades, particiones de grafos y agrupación. Con frecuencia, muchos algoritmos atraviesan el grafo de manera iterativa utilizando recorridos aleatorios, búsquedas de amplitud, búsquedas de profundidad o coincidencia de patrones. Debido al crecimiento exponencial de las posibles rutas al aumentar la distancia, muchos de los enfoques también tienen una alta complejidad

2 <https://neo4j.com/graph-data-science-library/>

algorítmica. Afortunadamente, existen algoritmos optimizados que utilizan ciertas estructuras del grafo, memorizan partes ya exploradas y paralelizan operaciones (Neo4j, 2020a).

Catálogo del grafo

Para ejecutar los algoritmos de la manera más eficiente posible, la biblioteca GDS utiliza la memoria RAM para representar los datos del grafo. Por lo tanto, es necesario cargar los datos en la memoria. La cantidad de datos cargados se puede controlar mediante las llamadas proyecciones de grafos, que también permiten, por ejemplo, filtrar por etiquetas de nodos y tipos de relaciones, entre otras opciones.

Modo de ejecución de algoritmos

Se deben tener en cuenta los modos de ejecución Stream, Stats, Mutate o Write, los cuales se describen a continuación:

- **Stream:** devuelve los resultados en memoria de la consulta ejecutada. Esto es similar a como funcionan las consultas estándares de Neo4j con su lenguaje Cypher.
- **Stats:** devuelve resultados estadísticos de la ejecución del algoritmo, como recuentos o distribuciones de percentiles. Este modo no realiza modificaciones o actualizaciones sobre los datos.
- **Mutate:** vuelve a escribir los resultados del cálculo del algoritmo en el grafo que está en memoria.
- **Write:** vuelve a escribir los resultados del cálculo del algoritmo en la base de datos de Neo4j. Este modo de ejecución fue el usado en los algoritmos aplicados a los datos de Mercado Libre.

Sintaxis

La sintaxis empleada para la ejecución de algoritmos de la librería GDS es la siguiente:

```
CALL gds.<algoritmo>.write({
  nodeProjection: <nombre del nodo>,
  relationshipProjection: {
    <relación>: {
      type: <relación>,
      orientation: <orientación>
    }
  },
  writeProperty: <nombre de la propiedad>
})
```

En la sintaxis anterior se deben asignar los valores respectivos al momento de ejecutar un algoritmo. El significado de los valores que se deben asignar en cada ejecución es:

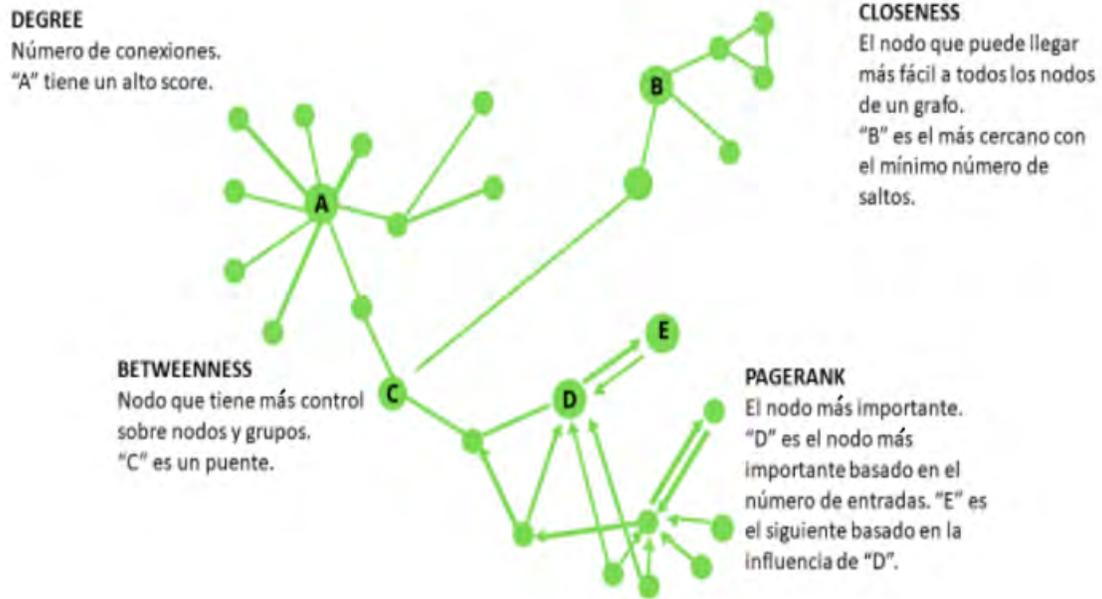
- **<algoritmo>:** nombre del algoritmo de la librería GDS.
- **<nombre del nodo>:** nombre del nodo sobre el que se va a aplicar el algoritmo.
- **<relación>:** relación que conecta a los nodos.
- **<orientación>:** orientación de la relación. Los valores permitidos para esta propiedad son **NATURAL** (se toma la orientación definida en el grafo), **UNDIRECTED** (no se tiene en cuenta la orientación de la relación) y **REVERSE** (se toma la dirección inversa definida en el grafo).
- **<nombre de la propiedad>:** campo donde queda guardado el resultado del algoritmo.

Los tipos de algoritmos que se pueden ejecutar en la librería GDS se describen a continuación.

Algoritmos de centralidad

Los algoritmos de centralidad se utilizan para descubrir los roles de nodos en un grafo y su impacto en la red. Estos identifican nodos importantes y ayudan a comprender la credibilidad, accesibilidad y velocidad a la que se propagan las cosas, así como los puentes entre los grupos. Estos algoritmos fueron creados para analizar redes sociales, pero han sido de utilidad en diferentes casos de uso. La figura 9 presenta algunos tipos de algoritmos de centralidad.

Figura 9. Algoritmos de centralidad



Fuente: elaboración propia.

Algoritmo de centralidad Degree

Se puede aplicar el algoritmo de centralidad *Degree* sobre las categorías de los productos para identificar cuál categoría tiene más conexiones u oferta mayor número de productos. El código fuente en Cypher aplicado es:

```
CALL gds.degree.write({
  nodeProjection: 'CATEGORIA',
  relationshipProjection: {
    CLASIFICADO_EN: {
      type: 'CLASIFICADO_EN',
      orientation: 'UNDIRECTED'
    }
  },
  writeProperty: 'DEGREE_SCORE'
})
```

La tabla 6 permite identificar que la categoría Samsung es la más importante según el número de conexiones que tiene con los demás nodos.

Tabla 6. Resultado del algoritmo de centralidad Degree sobre las categorías

Categoría	DEGREE_SCORE
Samsung	367
Xiaomi	287
Celulares y Teléfonos Celulares y Smartphones	235
Xiaomi Note 8 Pro	197
Xiaomi Note 9	182
Xiaomi Note 9S	171
Xiaomi Note 8	152
Apple iPhone 11	122
Huawei	113
Motorola	110

Fuente: elaboración propia.

Algoritmo de centralidad Weighted Degree

Este algoritmo nos ayuda a identificar el nodo más importante del grafo con base en el número de conexiones existentes, asignando un peso a cada conexión. Se aplica el algoritmo a las categorías, donde el peso de cada conexión es el número de productos vendidos. El código fuente en Cypher aplicado es:

```
CALL gds.degree.write({
  nodeProjection: 'CATEGORIA',
  relationshipProjection: {
    CLASIFICADO_EN: {
      type: 'CLASIFICADO_EN',
      orientation: 'UNDIRECTED',
      properties: 'CANTIDAD'
    }
  },
  relationshipWeightProperty: 'CANTIDAD',
  writeProperty: 'WEIGHT_DEGREE_SCORE'
})
```

En la tabla 7 se puede identificar que la categoría Xiaomi | Note8 es la más importante, según el número de conexiones con los demás nodos y la cantidad de productos vendidos.

Tabla 7. Resultado del algoritmo de centralidad *Weighted Degree* sobre las categorías

Código	WEIGHT_DEGREE_SCORE
Xiaomi Note 8	4247
Xiaomi Note 8 Pro	1444
Xiaomi	1099
Samsung A71 Duos	941
Apple iPhone 11	819
Xiaomi Note 9S	686
Xiaomi 7A	642
Xiaomi Note 9 Pro	611
Xiaomi Note 9	530
Xiaomi 8A	476

Fuente: elaboración propia.

Algoritmo de centralidad PageRank

Con este algoritmo se puede identificar la categoría más importante de acuerdo con el número de conexiones, tomando en cuenta la importancia de las categorías que la referencian. El código fuente en Cypher aplicado es:

```
CALL gds.pageRank.write({
  nodeProjection: "CATEGORIA",
  relationshipProjection: "CLASIFICADO_EN",
  maxIterations: 20,
  dampingFactor: 0.85,
  writeProperty: 'PAGE_RANK_SCORE'
})
```

Según la información en la tabla 8, las categorías “Celulares y Smartphones” y “Celulares y Teléfonos” tiene los dos puntajes más altos, debido a que son los nodos padres de todas las categorías y toman la importancia de las categorías hijas.

Tabla 8. Resultado del algoritmo de centralidad *PageRank* sobre las categorías

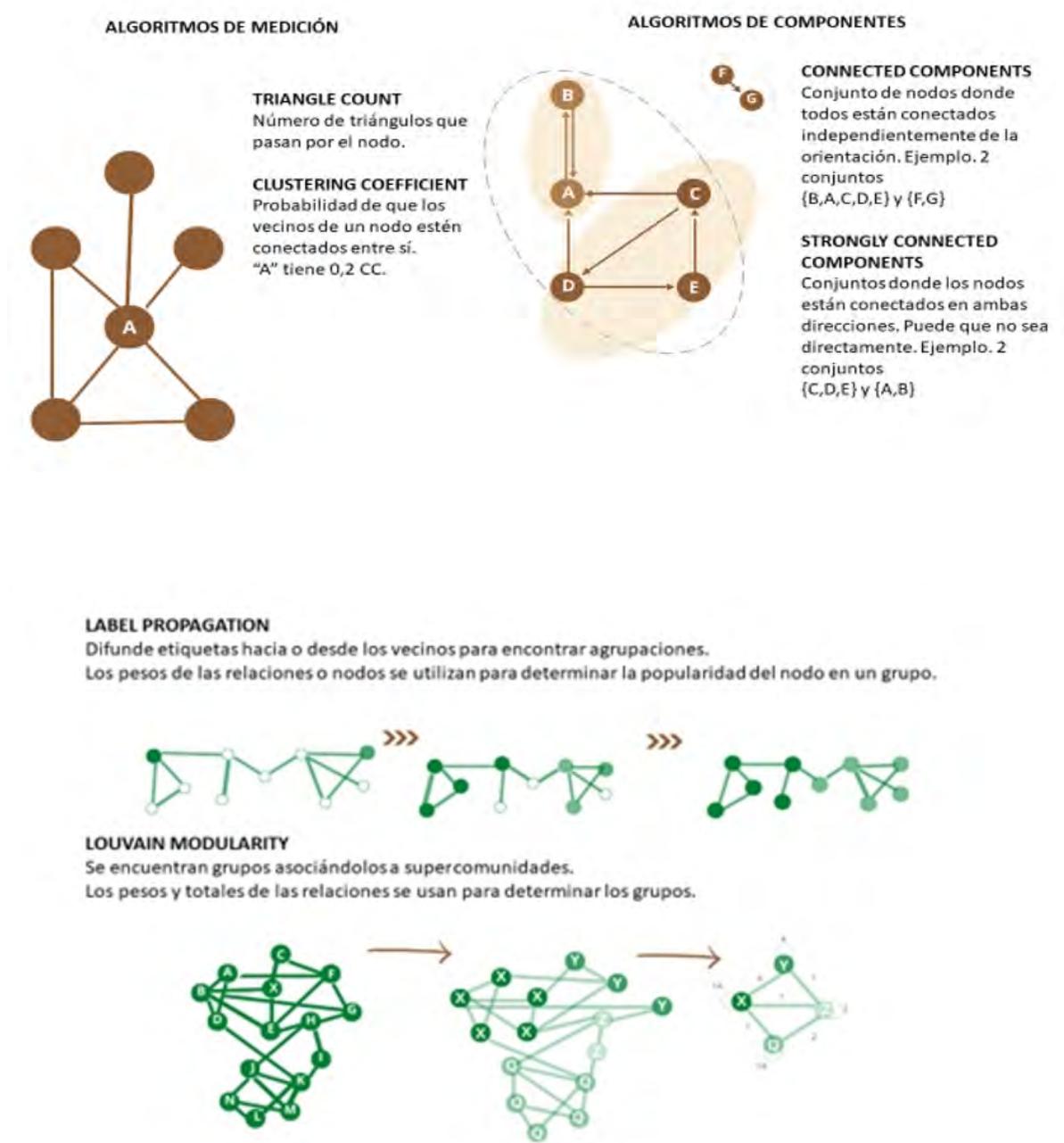
Código	PAGE_RANK_SCORE
Celulares y Smartphones	616
Celulares y Teléfonos	524
Samsung	197
Xiaomi	172
Apple	86

Fuente: elaboración propia.

Algoritmos de detección de comunidades

La formación de comunidades es frecuente en los esquemas de grafos debido a que esta ayuda a identificar el comportamiento grupal. Los miembros de cada comunidad tienen más relaciones dentro del grupo que con nodos fuera de él. Esto revela grupos de nodos, grupos aislados y la estructura de la red. Los algoritmos de detección de comunidades son usados para visualizar la red con fines de inspección general, como los expuestos en la figura 10.

Figura 10. Algoritmos de detección de comunidades



Fuente: elaboración propia.

Algoritmo de detección de comunidades *Connected Components*

Se aplica este algoritmo a los usuarios vendedores y compradores con el fin de identificar grupos de usuarios conectados entre sí. El código fuente en Cypher aplicado es:

```
CALL gds.wcc.write({
  nodeProjection: 'USUARIO',
  relationshipProjection: {
    OPINA: {
      type: 'OPINA',
      orientation: 'UNDIRECTED'
    }
  },
  writeProperty: 'WCC_COMUNIDAD'
})
```

Se identifica en la tabla 9 que existe una comunidad con 13.594 usuarios y 4 comunidades con muy pocos usuarios. Esto indica que la mayoría de los usuarios tiene el mismo comportamiento y las mismas conexiones.

Tabla 9. Resultado del algoritmo de detección de comunidades *Connected Components*

Comunidad	N.º usuarios
1	13.594
2	52
3	51
4	47
5	45

Fuente: elaboración propia.

Algoritmo de detección de comunidades *Label Propagation*

Se aplica este algoritmo a los usuarios vendedores y compradores para identificar grupos de usuarios que estén conectados entre sí usando las conexiones de los vecinos. El código fuente en Cypher aplicado es:

```
CALL gds.wcc.write({
  nodeProjection: 'USUARIO',
  relationshipProjection: {
    OPINA: {
      type: 'OPINA',
      orientation: 'UNDIRECTED'
    }
  },
  writeProperty: 'WCC_COMUNIDAD'
})
```

En la tabla 10 se identifica que este algoritmo distribuye de manera uniforme las comunidades de usuarios.

Tabla 10. Resultado del algoritmo de detección de comunidades *Label Propagation*

Comunidad	N.º usuarios
1	382
2	339
3	327
4	326
5	306

Fuente: elaboración propia.

Algoritmo de detección de comunidades *Louvain Modularity*

Se aplica este algoritmo a los usuarios vendedores y compradores para identificar supercomunidades. El código fuente en Cypher aplicado es:

```
CALL gds.labelPropagation.write({
  nodeProjection: 'USUARIO',
  relationshipProjection: {
    OPINA: {
      type: 'OPINA',
      orientation: 'UNDIRECTED'
    }
  },
  writeProperty: 'LP_COMUNIDAD'
})
```

En la tabla 11 se puede observar que este algoritmo distribuye de manera más uniforme las comunidades de usuarios, en comparación con los algoritmos anteriores.

Tabla 11. Resultado del algoritmo de detección de comunidades *Louvain Modularity*

Comunidad	N.º usuarios
1	393
2	341
3	339
4	337
5	328

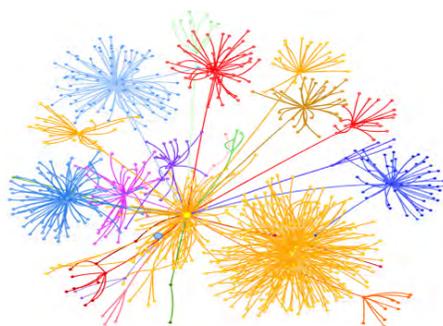
Fuente: elaboración propia.

Visualización de grafos

Usando los resultados de los algoritmos aplicados en las secciones anteriores se puede visualizar la importancia de los nodos y las comunidades identificadas. Para poder visualizar las estadísticas que proporcionan los algoritmos de Neo4j puede utilizarse la librería Neovis.js.³

Visualización de categorías de Mercado Libre

Usando las métricas obtenidas en los puntos anteriores, la figura 11 permite visualizar las comunidades identificadas para las categorías de productos de Mercado Libre, donde el tamaño del nodo indica la importancia dentro del grafo.

Figura 11. Visualización de comunidades de las categorías de Mercado Libre

Categoría	Apple	Incipio	
	Asus	Kyocera	
	BLU	LG	
	Google	Meizu	
	HTC	Nokia	
	Huawei	Samsung	

Fuente: elaboración propia.

Visualización de usuarios de Mercado Libre

Usando las métricas obtenidas en los puntos anteriores, la figura 12 indica que existen comunidades con una gran cantidad de usuarios pero pocas conexiones con otras comunidades. En contraste, hay comunidades más pequeñas que sí se relacionan con otras comunidades.

Figura 12. Visualización de comunidades de los usuarios de Mercado Libre

Fuente: elaboración propia.

CONCLUSIONES

Se recolectaron datos públicos de Mercado Libre para productos ofertados en la ciudad de Bogotá, aunque los usuarios vendedores no necesariamente están en la misma ciudad donde se ofertan tales productos, puesto que se identificó la presencia de vendedores provenientes de los departamentos de Atlántico, César y Cundinamarca. Además, Bogotá está registrado en el sistema como un departamento, mientras que sus localidades figuran como municipios, siendo Suba aquel con más usuarios vendedores.

Gracias a la notación intuitiva de nodos y relaciones, el lenguaje Cypher es mucho más sencillo que SQL; más aún si se aumenta el número de *joins* o un patrón de búsqueda. Además, el lenguaje SQL resulta ser menos eficiente, al tener que recorrer todas las tablas mediante *joins*.

Con el método empleado se pueden detectar los casos de baja oferta y alta demanda para un producto, como sucede con el dispositivo Xiaomi Note 8, que resultó ser el producto más vendido (figura 2), a pesar de que ocupa el sexto lugar dentro de los productos más ofertados (figura 4).

El usuario vendedor “Celulares 99” ingresó a la plataforma el 1 de agosto de 2020 y ha vendido 1.047

3 <https://github.com/neo4j-contrib/neovis.js/>

productos. Por su parte, el usuario vendedor “Descobar78” ingresó el 1 de septiembre de 2007 y ha logrado vender 446 productos. Haciendo esta comparación, “Celulares 99” es un usuario con un buen perfil de ventas para Mercado Libre, debido a que en pocos meses ha logrado alcanzar ventas que otros usuarios no han alcanzado en años de operación.

Las consultas en lenguaje declarativo Cypher revelan información de cantidades de productos, vendedores, compradores y categorías, mientras que los algoritmos de analítica de grafos son adecuados para revelar patrones de la relación entre vendedores y compradores sin estar directamente conectados. Por ello, la visualización de grafos contribuye a dar sentido al volumen de usuarios y las categorías de productos, además de identificar comunidades que no son visibles a simple vista.

La información esquematizada debe tener características puntuales para que un algoritmo de grafos permita obtener un buen resultado. Esto se pudo observar al aplicar el algoritmo de detección de comunidades *Connected Components*, puesto que no se distribuyó la información en comunidades homogéneas, dando como resultado que la mayoría de los usuarios fueran clasificados dentro de una misma comunidad.

En conclusión, la analítica de grafos aplicada a los datos de Mercado Libre ayuda a identificar las categorías ofertadas más importantes, así como los usuarios más influyentes dentro del sistema. De igual forma, la aplicación de este método hace que se pueda categorizar la información en comunidades, tomando como criterio características similares que no se identifican a simple vista.

Referencias

- Branting, L. K., Reeder, F., Gold, J., & Champney, T. (2016). Graph analytics for healthcare fraud risk estimation. *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 845-851). ASONAM. <https://www.computer.org/csdl/proceedings-article/asonam/2016/07752336/12OmNz4SOvm>
- Carrillo-Gelvez, G. (2019). *Conjunto de datos de Mercadolibre* [data set]. GitHub. https://github.com/gersongelvez/TESIS_MAESTRIA/tree/master/DATOS
- Das, S. R., & Sisk, J. (2005). Financial communities. *Journal of Portfolio Management*, 31(4), 112-123.
- DB-Engines (2020a). *DB-Engines ranking*. <https://db-engines.com/en/ranking>
- DB-Engines (2020b). *DB-Engines ranking of graph DBMS*. <https://db-engines.com/en/ranking/graph+dbms>
- Dinero. (2020, agosto 10). *Mercado Libre: ¿cómo llegó a ser la firma más valiosa de Latinoamérica?* <https://www.dinero.com/empresas/articulo/mercado-libre-es-la-empresa-mas-valiosa-de-america-latina-en-2020/295269>
- Eboli, M. (2007). *Systemic risk in financial networks: A graph-theoretic approach*. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.466.7515&rep=rep1&type=pdf>
- Kanavos, A., Drakopoulos, G., & Tsakalidis, A. (2017). Graph community discovery algorithms in Neo4j with a regularization-based evaluation metric. *Proceedings of the 13th International Conference on Web Information Systems and Technologies (WEBIST 2017)* (pp. 403-410). WEBIST. <https://www.scitepress.org/papers/2017/63821/63821.pdf>
- Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5), 604-632. <https://doi.org/10.1145/324133.324140>
- Molloy, I., Chari, S., Finkler, U., Wiggerman, M., Jonker, C., Habeck, T., Park, Y., Jordens, F., & van-Schaik, R. (2017). Graph analytics for real-time scoring of cross-channel transactional fraud. En J. Grossklags & B. Preneel (eds.), *Financial Cryptography and Data Security* (pp. 22-40). Springer. https://doi.org/10.1007/978-3-662-54970-4_2
- Neo4j. (2020a). *Neo4j Graph Data Science Library*. <https://neo4j.com/graph-data-science-library/>

- Neo4j. (2020b). *The Graph of Thrones [Season 7 Contest]*. <https://neo4j.com/blog/graph-of-thrones/>
- Neo4j (2020c). *Neo4j Graph database concepts*. <https://neo4j.com/docs/getting-started/current/graphdb-concepts/#graphdb-concepts>
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank Citation Ranking: Bringing order to the web*. Stanford InfoLab. <http://ilpubs.stanford.edu:8090/422/>
- Pragma. (2018). *Los beneficios de las bases de datos nosQL*. <https://www.pragma.com.co/blog/los-beneficios-de-las-bases-de-datos-nosql>
- Ribeiro, J., Silva, P., Duarte, R., Davids, K., & Garganta, J. (2017). Team sports performance analysed through the lens of social network theory: Implications for research and practice. *Sports Medicine*, 47(9), 1689-1696.
- Rossi, R. A., & Ahmed, N. K. (2015). *The network data repository with interactive graph analytics and visualization*. *arXiv*, 2. <http://arxiv.org/abs/1410.3560>
- Sadowski, G., & Rathle, P. (2014). *Fraud detection: Discovering connections with graph databases*. Neo4j. <https://neo4j.com/whitepapers/white-paper-fraud-detection/>
- Scott, J., & Carrington, P. (2011). *The SAGE handbook of social network analysis*. SAGE.
- Scott, J. (2011). Social network analysis: Developments, advances, and prospects. *SOCNET*, 1, 21-26. <https://doi.org/10.1007/s13278-010-0012-6>
- Svensson, J. (2020). *SDTimes*. <https://sdtimes.com/databases/guest-view-relational-vs-graph-databases-use/>
- Pérez, J., Arenas, M., & Gutierrez, C. (2009). Semantics and complexity of SPARQL. *ACM Transactions on Database Systems*, 34(3), 1-45.
- Vicknair, C., Macias, M., Zhao, Z., Nan, X., Chen, Y., & Wilkins, D. (2010). A comparison of a graph database and a relational database. *Proceedings of the 48th Annual Southeast Regional Conference* (pp. 1-6). ACM. <https://doi.org/10.1145/1900008.1900067>
- Virji-Babul, N., Hilderman, C. G. E., Makan, N., Liu, A., Smith-Forrester, J., Franks, C., & Wang, Z. J. (2014). Changes in functional brain networks following sports-related concussion in adolescents. *Journal of Neurotrauma*, 31(23), 1914-1919.
- Weng, J., Lim, P., Jiang, J., & He, Q. (2010). Twitterrank: Finding topic-sensitive influential Twitterers. *Proceedings of the third ACM International Conference on Web Search and Data Mining* (261-270). ACM.

MUTIS

REVISTA DE CIENCIAS NATURALES E INGENIERÍA DE LA UNIVERSIDAD JORGE TADEO LOZANO

Instrucciones para autores

ENVÍOS DE ARTÍCULOS

La revista *Mutis* acepta permanentemente los siguientes tipos de artículos:

1. Artículo de investigación científica y tecnológica: documento que presenta, de manera detallada, los resultados originales de proyectos terminados de investigación.
1. Artículo de métodos: documento que presenta la invención o estandarización de un método experimental, computacional, o educativo, entre otros.
1. Artículo de revisión (estado del arte): es una revisión bibliográfica de por lo menos 50 referencias, con el fin de dar cuenta sobre los avances y las tendencias de desarrollo de un campo particular de las ciencias naturales e ingeniería.
1. Comunicación corta: documento que presenta resultados parciales de una investigación en desarrollo cuya publicación sea de gran importancia.

Los trabajos que se presentan a consideración de la revista *Mutis*, se deben enviar a través de nuestro sistema de gestión editorial <http://goo.gl/m6Q-6FD> o vía correo electrónico a revista.mutis@utadeo.edu.co (Asunto: artículo revista *Mutis*) en formato de procesador de texto habilitado para modificaciones (Microsoft Word). El artículo debe estar escrito en español o inglés, con una extensión máxima de 25 páginas (incluyendo tablas, figuras y referencias), con márgenes simétricas de 2 cm, sin columnas, con títulos y subtítulos alineados a la izquierda, en letra Times New Roman, tamaño 12 y en espacio doble. Para facilitar el proceso de evaluación, se deben numerar todos los renglones del trabajo. Los trabajos deben tener la siguiente estructura y cumplir con los siguientes requisitos:

Título. El título debe representar con precisión y de manera concisa el contenido del trabajo sin exceder 15 palabras. Se debe escribir con letras tamaño de fuente 16, en minúsculas, en negrilla y alineado a la izquierda, en idioma original y en inglés.

Autores. Debe escribirse nombre completo de los autores, separados por comas y tamaño de fuente 12. No se deben incluir títulos académicos ni las posiciones o cargos que los autores ocupan. Las afiliaciones se deben escribir en este orden: universidad, facultad, programa, laboratorio o grupo (si aplica), dirección, ciudad, país, correo electrónico del autor, número de registro en ORCID y Google Scholar. Ejemplo:

Yvonne Sarmiento Rangel,¹ Ariadna Hazel Vergel^{2*}

¹**Universidad Jorge** Tadeo Lozano, Facultad de Ciencias Naturales e Ingeniería, Departamento de Ciencias Naturales y Ambientales. Grupo de Investigación Genética, Biología Molecular y Bioinformática – (GENBIMOL), Carrera 4 n.º 22-61, Bogotá D. C., Colombia. CP 011110

²**Instituto de** Investigaciones de Viandas Tropicales (INIVIT), Apartado 6, Santo Domingo, Villa Clara, Cuba. CP 53000. Correo electrónico: arhazel@inivit.cu <https://orcid.org/xxx>

https://scholar.google.es/citations?hl=es&user=pTZo5RoAAAJ&view_op=list_works

Resumen. Debe tener un máximo de 250 palabras. El texto del resumen debe ir en tamaño de fuente 12 Times New Roman. El resumen debe incluir objetivo, métodos, resultados concretos de la investigación y una conclusión general.

Palabras clave. Deben escogerse un máximo de 5 palabras claves. Palabras diferentes incluidas en el título.

Abstract. Debe ser una traducción correcta y precisa al idioma inglés del resumen que aparece en español. Se usa el mismo estilo ya descrito para el resumen en español.

Keywords. Debe ser una traducción correcta y precisa al idioma inglés de las palabras clave que aparecen en español. Se usa el mismo estilo ya descrito para las palabras clave en español.

Introducción. Donde se habla sobre el origen de la investigación, interés y visión general del tema a tratar.

Materiales y métodos. No aplica para artículo de revisión. Descripción detallada de técnicas, equipos y reactivos utilizados en el desarrollo de la investigación.

Resultados. No aplica para artículo de revisión. Donde se hace una presentación clara de los resultados de investigación.

Discusión. Donde se realiza la interpretación y análisis de los resultados de investigación comparándolos con resultados previamente publicados. Pueden presentarse tablas o figuras para sintetizar los resultados. Si se han realizado análisis estadísticos, se deberá describir la procedencia de los datos y el método estadístico empleado.

Conclusiones. Deben basarse en los resultados obtenidos y estar relacionadas con la hipótesis o planteamiento del problema descrito en la introducción.

Ecuaciones. Las ecuaciones se deben escribir centradas con respecto al texto principal y se deben numerar continuamente a lo largo del texto. Cada ecuación debe estar enumerada con un número entre paréntesis ubicado en el margen derecho del texto. El significado y las unidades deben explicarse inmediatamente después de la ecuación en que se utilicen por primera vez.

Cantidades y unidades. Para todas las cantidades numéricas, se debe utilizar una coma para separar los miles y un punto para indicar las cifras decimales. Se recomienda utilizar el Sistema Internacional de Unidades.

Tablas. Las tablas se deben numerar continuamente a lo largo del texto. El encabezado de cada tabla debe incluir la palabra Tabla (en negrilla) seguida del número arábigo consecutivo correspondiente y de una breve descripción del contenido de la tabla. Tanto el

encabezado como el contenido de la tabla se deben escribir en tamaño de fuente 10. Solo deben usarse líneas horizontales para separar las entradas de la tabla. No deben usarse líneas verticales para separar las columnas. En el caso de necesitarse algún pie de tabla, este debe estar escrito con tamaño de fuente 8. Las tablas no deben duplicar la información dada en las figuras.

Figuras. Las figuras (o fotografías) se deben numerar de manera consecutiva a lo largo del texto. El rótulo de cada figura debe estar en la parte inferior de la figura y debe incluir la palabra Figura (en negrilla) seguida del número arábigo correspondiente y de una breve descripción del contenido de la figura. El rótulo de la figura se debe escribir en tamaño de fuente 10. Las figuras deben incluirse como imágenes originales importadas o copiadas al archivo del trabajo desde una herramienta gráfica (por ejemplo, *Microsoft Excel*) con una resolución de 300 dpi, de tal manera que la legibilidad no se sacrifique cuando el tamaño de las figuras se ajuste durante el proceso de edición. No se publicará por ningún motivo información de fotografías descargadas de internet sin el permiso expreso del autor.

Referencias bibliográficas. Se debe utilizar a lo largo del texto el sistema de citación APA que utiliza los apellidos de autores seguidos del año de publicación entre paréntesis. La lista de las referencias se debe dar al final del documento después de la sección de Conclusiones y debe estar organizada en orden alfabético de acuerdo a la inicial del primer apellido del primer autor. En esta lista solamente se deben colocar las referencias citadas en el texto. Todas las referencias se deben escribir en tamaño de fuente 12 Times New Roman.

EJEMPLOS:

ARTÍCULO

Giraldo, G., Talens, P., Fito, P., & Chiralt, A. (2003). Influence of sucrose solution concentration on kinetics and yield during osmotic dehydration of mango. *Journal of food Engineering*, 58(1), 33-43.

LIBRO

Shames, I. H., & Shames, I. H. (1982). *Mechanics of fluids* McGraw-Hill.

CAPÍTULO DE LIBRO

Bertucco, A. &. (2002). Micronization of a polysaccharide by a supercritical antisolvent technique. En J. &. Williams (ed.), *Supercritical Fluids Methods and Protocols* (pp. 335-412). Humana Press, Inc.

DOCUMENTO DE INTERNET

Benson, T. (Nov. 17, 2005). *Forces on a Rocket*. [http:// exploration.grc.nasa.gov/education/rocket/rktfor.html](http://exploration.grc.nasa.gov/education/rocket/rktfor.html), accesado el XX de XX de XXXX

TESIS DE GRADO

Yépez, B. (2009). *Glicerólisis e hidrólisis de triglicéridos saturados, en dióxido de carbono supercrítico* (tesis doctoral). Escuela de Ingeniería Química, Universidad del Valle, Cali, Colombia.

En caso de dudas, comunicarse con el editor (revista.mutis@utadeo.edu.co). Estas instrucciones quedan supeditadas a los cambios que el Comité Editorial disponga.



ESTA REVISTA ELECTRÓNICA
ES EDITADA POR LA
UNIVERSIDAD DE BOGOTÁ
JORGE TADEO LOZANO.
SE FINALIZÓ EN 2021