



**Desarrollo de modelo de medición de riesgo de lavado de activos y financiación de
terrorismo LAFT en las jurisdicciones nacionales.**

Karla Andrea Molina Echeverry

Informe trabajo de grado presentado para optar al Título de Ingeniera Industrial

Asesor

Miguel Ángel Arroyave Guerrero, M. Sc. Ingeniería

Universidad de Antioquia

Facultad de Ingeniería

Ingeniería Industrial

Medellín, Colombia

2023

Cita	(Molina Echeverry, 2024)
Referencia	Molina Echeverry K.A (2024). <i>Desarrollo de modelo de medición de riesgo de lavado de activos y financiación del terrorismo en las jurisdicciones nacionales</i> [Proyecto práctica profesional]. Universidad de Antioquia, Medellín.
Estilo APA 7 (2020)	



Créditos a escenario de prácticas, personas, proyectos que aportaron al desarrollo de la práctica (interna y externamente: empresa y área de la empresa, grupo de investigación, proyecto, organización)



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano/Director: Julio César Saldarriaga Molina.

Jefe departamento: Mario Alberto Gaviria Giraldo.

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Agradecimientos

En primer lugar, deseo expresar mi profundo agradecimiento a mi familia y compañeros, por su inquebrantable respaldo emocional y motivación constante. Sus palabras alentadoras y su apoyo incondicional fueron un pilar fundamental para superar los desafíos que surgieron en este proceso. Al Grupo Bancolombia por brindarme la oportunidad de realizar mis prácticas en su empresa y especialmente a la Gerencia de Metodologías de Riesgos de Cumplimiento, su cálida bienvenida, amabilidad y disposición para compartir sus conocimientos han hecho de este ambiente laboral un lugar enriquecedor y sobre todo acogedor; Agradezco a cada uno por su colaboración y apoyo, por hacer de mi tiempo en la organización una experiencia inolvidable. Gracias por ser parte fundamental de este importante capítulo en mi vida académica.

Tabla de contenido

Resumen	8
Abstract	9
Introducción	10
1.Objetivos	12
1.1 Objetivo general	12
1.2 Objetivos específicos	12
2 Marco teórico	13
3 Metodología	20
4 Resultados y análisis	22
4.1 Indagación de la información	22
4.2 Modelado	27
4.3 Herramienta de usuario final	33
5. Conclusiones	39
6. Recomendaciones	41
7.Referencias	42

Lista de tablas

Tabla 1 Variables Seleccionadas	23
Tabla 2 Importancia de las variables por categoria	32

Lista de figuras

Figura 1 Principios para el desarrollo del SARLAF en Bancolombia	15
Figura 2 Distribución de las variables	26
Figura 3 Diagrama de correlación	27
Figura 4 Reporte de métricas Gradient Boostier Classifier	29
Figura 5 Matriz de Confusión Gradient Boostier Classifier	30
Figura 6 Curva ROC Gradient Boostier Classifier	30
Figura 7 Variables Importantes Modelo Gradient Boostier Classifier	31
Figura 8 Pantalla inicio del tablero PBI	33
Figura 9 Pantalla del tablero PBI con municipio seleccionado	34
Figura 10 Pantalla calificación del tablero PBI	35
Figura 11 Pantalla detallad de datos del tablero PBI	36
Figura 12 Pantalla descarga del tablero PBI	37
Figura 13 Pantalla descripción de las variables del tablero PBI	38

Siglas, acrónimos y abreviaturas

LAFT	Lavado de activos y financiación del terrorismo
ROC	Receiver Operating Characteristic curve
AUC	Área Under the Curve
SARLAFT	Sistema de administración de riesgo de lavado de activos y financiación del
terrorismo	

Resumen

La gestión eficaz del riesgo de lavado de activos y financiación del terrorismo (LAFT) es esencial para cualquier organización, ya que permite abordar amenazas y vulnerabilidades y tomar medidas preventivas para mitigar su impacto negativo. En este contexto, se implementó el manual SARLAFT 4.0, que comprende las etapas de identificación, medición, control y monitoreo en el ciclo de gestión de riesgo; En el proceso de medición es necesario medir el riesgo que tiene el grupo Bancolombia en cada uno de los factores los cuales son: clientes, canales, productos y jurisdicciones.

Para medir el riesgo que tienen las jurisdicciones se realizó una herramienta de usuario final de calificación de riesgo, donde por medio de selección de variables y recolección de información se desarrolló un modelo de machine learning usando un algoritmo de aprendizaje supervisado con el cual determinamos el nivel de riesgo que tienen las jurisdicciones, donde finalmente el modelo identificó 205 municipios con riesgo alto, 179 con riesgo medio y 737 con riesgo bajo. Permitiéndole a la organización tomar decisiones más ágiles y efectivas frente a posibles amenazas con un enfoque basado en riesgos buscando la mitigación del riesgo LAFT en las jurisdicciones, contribuyendo así a la estabilidad económica y al desarrollo sostenible tanto a nivel organizacional como en el país.

Palabras clave: SARLAF, LAFT, riesgo, jurisdicciones, machine learning, modelo, score.

Abstract

The effective management of money laundering and terrorist financing (ML/TF) risk is essential for any organization, as it enables addressing threats and vulnerabilities and taking preventive measures to mitigate their negative impact. In this context, the SARLAFT 4.0 manual was implemented, comprising the stages of identification, measurement, control, and monitoring in the risk management cycle. In the measurement process, it is necessary to assess the risk that Bancolombia Group has in each of the factors, which are: clients, channels, products, and jurisdictions.

To measure the risk associated with jurisdictions, a user-end risk assessment tool was developed. Through the selection of variables and information gathering, a machine learning model was created using a supervised learning algorithm. This model determined the level of risk for each jurisdiction, ultimately identifying 205 municipalities with high risk, 179 with medium risk, and 737 with low risk. This approach enables the organization to make more agile and effective decisions in response to potential threats, with a risk-based focus aimed at mitigating ML/TF risk in jurisdictions.

This contributes to economic stability and sustainable development at both the organizational and national levels. The risk mitigation efforts are aligned with a risk-based approach, fostering economic stability and sustainable development in the organization and the country.

Keywords: SARLAF, LAFT, risk, jurisdictions, machine learning, model, score.

Introducción

El Grupo Bancolombia actúa de manera consistente con su propósito de promover el desarrollo sostenible para el bienestar de todos, y está comprometido en la lucha contra los riesgos que tienen asociados.

El riesgo de lavado de activos y financiación del terrorismo es uno de ellos y representa una amenaza a la estabilidad del sistema económico, es un obstáculo para el desarrollo sostenible y el crecimiento financiero tanto de las organizaciones públicas como privadas del país, por lo que hacer una correcta gestión de este riesgo buscando evitar que este ocurra es fundamental para la organización.

En el ciclo de gestión de riesgos se utilizan herramientas como el manual SARLAFT 4.0, el cual se enfoca en promover un enfoque basado en riesgo, por lo que para el desarrollo de este en el Grupo Bancolombia se han creado políticas, procedimientos, metodologías y herramientas para la identificación, medición, control y monitoreo del riesgo LAFT.

Una parte fundamental para poder realizar una correcta gestión de riesgo es identificar este dónde se presenta y como se lleva a cabo, teniendo en cuenta esta premisa el grupo Bancolombia realiza esta identificación basándose en cuatro factores de riesgo: clientes, productos, canales y jurisdicciones; Para esto el grupo Bancolombia cuenta con herramientas como los Score de medición de riesgos asociados a cada uno de los factores.

El score de jurisdicciones nacionales es una de esas herramientas, este es fundamental para realizar la identificación de las jurisdicciones más riesgosas en lavado de activos y financiación del terrorismo. El grupo Bancolombia hace presencia en cerca del 98.9% de las

DESARROLLO DE MODELO DE MEDICIÓN DE RIESGO DE LAVADO DE ACTIVOS Y FINANCIACIÓN DEL TERRORISMO EN LAS JURISDICCIONES NACIONALES.

jurisdicciones nacionales por lo que debe poder generar controles para evitar y mitigar este tipo de riesgo.

Actualmente para realizar este score el banco cuenta con un proveedor el cual es el encargado de realizar la actualización de este, pero se ha evidenciado que desde el 2021 no se ha actualizado, y en vista de los cambios constantes que se tienen a nivel nacional en las jurisdicciones esta calificación debe ser actualizada año a año, ya que, sin una correcta actualización el banco no podrá tomar medidas de prevención y mitigación de riesgo efectivas.

Por lo tanto, con este proyecto se creó un nuevo score de jurisdicciones nacionales por medio de un modelo analítico que nos permitió calificar las jurisdicciones nacionales en alto, medio, bajo permitiéndonos medir el riesgo LAFT en estas. Este modelo identificó 205 municipios con riesgo alto, 179 con riesgo medio y 737 con riesgo bajo.

Este enfoque proporciono al Grupo Bancolombia una herramienta valiosa para tomar decisiones informadas y estratégicas en la prevención y mitigación del riesgo LAFT a nivel nacional, además de no solo contribuir a la seguridad financiera de la organización, sino que también demuestra su compromiso continuo con el desarrollo sostenible y el bienestar de la sociedad.

1.Objetivos

1.1 Objetivo general

Desarrollar un modelo analítico que permita determinar el nivel de riesgo de lavado de activos y financiación del terrorismo (LAFT) para cada una de las jurisdicciones nacionales, con el propósito de permitir a la organización tomar decisiones fundamentadas y estratégicas para mitigar y prevenir este riesgo, contribuyendo así a la estabilidad económica y al desarrollo sostenible tanto a nivel organizacional como en el país.

1.2 Objetivos específicos

- Analizar y seleccionar las variables relevantes en cada jurisdicción, basadas en los delitos fuente de lavado de activos, para evaluar el riesgo de LAFT, y recopilar esta información de fuentes gubernamentales y bases de datos internas de la organización.
- Desarrollar un modelo analítico que permita calificar el nivel de riesgo de LAFT en cada jurisdicción, con el fin de facilitar la toma de decisiones estratégicas en la gestión de riesgos LAFT a través de la utilización de los resultados obtenidos.
- Validar el modelo analítico mediante pruebas y análisis de sensibilidad buscando garantizar su precisión y su robustez en la medición del riesgo LAFT.
- Generar una herramienta de usuario final con los resultados y análisis de resultados basados en el modelo analítico para proporcionar a la organización información actualizada sobre el riesgo de LAFT en cada jurisdicción.

2 Marco teórico

Las organizaciones financieras tienen un riesgo permanente que es el de lavado de activos y financiación del terrorismo LA/FT, y es su deber como organización evitar o mitigar que este riesgo se materialice.

Se entiende por riesgo de lavado de activos y de la Financiación del terrorismo LA/FT la posibilidad de pérdida o daño que puede sufrir una entidad vigilada por su propensión a ser utilizada directamente o a través de sus operaciones como instrumento para el lavado de activos y/o canalización de recursos hacia la realización de actividades terroristas, o cuando se pretenda el ocultamiento de activos provenientes de dichas actividades. (Colombia B. d., 2014)

El lavado de activos y la financiación del terrorismo representan una gran amenaza para la estabilidad del sistema financiero y la integridad de los mercados por su carácter global y las redes utilizadas para el manejo de tales recursos. Tal circunstancia destaca la importancia y urgencia de combatirlos. (Colombia S. f., 2022)

Las organizaciones financieras están regidas bajo normativas que regulan y supervisan las operación y gestiones que se realizan día a día en estas, la Superintendencia Financiera de Colombia tiene por objetivo supervisar el sistema financiero colombiano con el fin de preservar su estabilidad, seguridad y confianza, así como promover, organizar y desarrollar el mercado de valores colombiano y la protección de los inversionistas, ahorradores y asegurados, este es el ente supervisor y regulador de las entidades financieras además, esta impuso la adopción del Sistema de administración del riesgo de lavado de activos y de la Financiación del terrorismo SARLAF por parte de las entidades vigiladas, las cuales están en la obligación de prevenir la posibilidad de pérdida o daño que pueden sufrir por su propensión a ser utilizadas directamente por sus

accionistas, administradores o funcionarios o a través de sus operaciones como instrumento para el lavado de activos y/o canalización de recursos hacia la realización de actividades terroristas, por sus clientes o usuarios (Colombia S. f., 2022).

El sistema de administración de riesgo de lavado de activos y financiación del riesgo SARLAF 4.0 está dentro del Capítulo IV del Título IV en la Parte I de la Circular Básica Jurídica de la Superintendencia Financiera de Colombia, esta es una norma que da a las organizaciones unas pautas a las organizaciones para la implementación de medidas para identificar, evaluar y mitigar los riesgos asociados al LAFT, cada organización es la encargada de realizar el desarrollo del sistema acorde a sus necesidades y objetivos organizacionales (Colombia S. f., 2022).

El SARLAFT 4.0 se desarrolla en cuatro etapas:

1. Identificación del riesgo de LA/FT: En esta primera etapa las entidades deben establecer metodologías para segmentar los factores de riesgo, teniendo en cuenta el contexto interno y externo de la entidad con el fin de identificar las formas a través de las cuales se puede presentar un riesgo considerando todas las variables. Estos riesgos identificados se conocen como riesgo inherente.
2. Medición de la probabilidad y el impacto del riesgo de LA/FT: En esta segunda etapa se deben establecer metodologías para medir la probabilidad de ocurrencia y el impacto que esta tendría en caso de llegar a materializarse, con el fin de crear un perfil de riesgo de la organización.
3. Control del riesgo de LA/FT: En la etapa de control de deben establecer medidas de control y establecer una evaluación de estas para analizar si el control está siendo efectivo. Estos controles deben reducir la probabilidad de ocurrencia o el impacto que pueda generar al materializarse el

riesgo, en caso de no ser así el control no está siendo efectivo. El riesgo después de aplicarle controles se conoce como riesgo residual.

4. Monitoreo del riesgo de LA/FT: En esta etapa se desarrolla un procedimiento de seguimiento efectivo donde se detecten y/o corrijan deficiencias.

Para el desarrollo del SARLAFT en el Grupo Bancolombia se han creado políticas, procedimientos, metodologías y herramientas para la identificación, medición, control y monitoreo del riesgo, buscando que el mismo incluya los componentes para el diseño y desarrollo de sus principios como los veremos en la figura 1.

Figura 1.

Principios para el desarrollo del SARLAF en Bancolombia



Nota Fuente: Bancolombia (2023)

La gestión del riesgo se mejora continuamente a través del aprendizaje y la experiencia, para lo cual las entidades tienen como deber contar con fuentes de información actualizada que

contengan cambios normativos en materia de LA/FT y provean los últimos acontecimientos en este sentido.

Por lo tanto, las instituciones financieras/servicios para dar cumplimiento a esto se han valido de diversas herramientas/métodos, entre los que se encuentra se apoya en herramientas que permitan la automatización de procesos buscando ser más efectivos y más precisos, por ello disciplinas como el Machine Learning que actualmente son usadas en la organización para muchos procesos como lo son servicios de créditos y consumo, conocimiento del cliente, autorizaciones y tramites, detección de irregularidades también puede usarse con el fin de identificar riesgos (Castro, 2022).

El Machine Learning (aprendizaje automático) es una disciplina de las ciencias informáticas relacionada con el desarrollo de inteligencia artificial, así como también con el análisis predictivo o aprendizaje estadístico por ello es de gran utilidad en la organización y esto básicamente hace referencia a la capacidad de que los sistemas pueden aprender de los datos, identificar patrones y tomar decisiones con una mínima intervención humana aprendiendo de unos datos ingresados al sistema (Mustafa , 2019)

El machine Learning utiliza conjuntos de instrucciones de modelado matemático que se denominan algoritmos y el conjunto de datos que alimenta el algoritmo se organiza en una serie de variables de entrada y tendrá como objetivo encajar en una o varias variables de salida. En base en estas variables de salida, se define lo que puede ser un diagnóstico o resultados relevantes (Mustafa , 2019).

El proceso de aprendizaje en el que se ajustan los parámetros matemáticos del modelo se denomina entrenamiento y la evaluación del modelo ajustado se denomina validación. (Cánovas, Cubo, Mir Rivera , & Ferro, 2023).

Para entrenar un modelo de Machine Learning es importante dividir el conjunto de datos en dos conjuntos de datos más pequeños que serán utilizadas de dos maneras: entrenamiento y prueba.

El subconjunto de datos de entrenamiento es empleado para estimar los parámetros del modelo y el subconjunto de datos de prueba se utiliza para evidenciar el comportamiento del modelo estimado (Mustafa , 2019).

Cada registro de la base de datos debe aparecer en uno de los dos subconjuntos, y para dividir el conjunto de datos en ambos subconjuntos, se utiliza un procedimiento de muestreo: muestreo aleatorio simple (Parra, 2019).

El machine Learning tiene una rama conocida como aprendizaje supervisado este se da cuando se tiene variables de entrada (x) y una variable de salida (Y) y se utiliza un algoritmo para aprender la función de mapeo (Patrones) de la entrada a la salida, el objetivo es aproximar la función de mapeo tan bien que cuando se tengan nuevos datos de entrada (x) se pueda predecir la variable de salida (Y) para esos datos (Parra, 2019).

Se llama aprendizaje supervisado porque el proceso de aprender de un algoritmo de unos datos puede ser visto como que un profesor está supervisando el proceso. Se conocen las respuestas correctas, así que el algoritmo va realizando predicciones de manera iterativa con los datos que tiene y, estas predicciones, son corregidas por el profesor, el aprendizaje termina cuando el algoritmo consigue obtener unos resultados aceptables. (Marrugat, 2020)

Los problemas de aprendizaje supervisados pueden agruparse en problemas de regresión y clasificación.

- Clasificación: Un problema de clasificación es cuando la variable de salida es una categoría.
- Regresión: Un problema de regresión es cuando la variable de salida es un valor real.

La clasificación es un proceso de dos pasos, paso de aprendizaje y paso de predicción. En el paso de aprendizaje, el modelo se desarrolla en base a datos de capacitación dados. En el paso de predicción, el modelo se usa para predecir la respuesta para datos dados (Castro, 2022).

En tareas de clasificación existen diversos algoritmos que son de gran utilidad cuando se requiere que la salida sea una categoría, como lo son los árboles de decisión estos son algoritmos de aprendizaje basados en árboles se consideran uno de los mejores y más utilizados métodos de aprendizaje supervisado. Los métodos basados en árboles potencian los modelos predictivos con alta precisión, estabilidad y facilidad de interpretación (Castro, 2022).

Un árbol de decisión en Machine Learning es una estructura de árbol similar a un diagrama de flujo donde un nodo interno representa una característica (o atributo), la rama representa una regla de decisión y cada nodo hoja representa el resultado. Esta técnica es la más básica dentro de la familia de los árboles, el algoritmo no es paramétrico y puede manejar de manera eficiente conjuntos de datos grandes y complicados sin imponer una estructura paramétrica complicada. (Piñeiro & Nakab, 2022)

El concepto fundamental detrás del Random Forest es que son una modificación sustancial del bagging que construye una gran colección de árboles descorrelacionados y luego los promedia. En muchos problemas, el rendimiento de los bosques aleatorios es muy similar al de los árboles que utilizas boosting, pero con la ventaja que son más sencillos de entrenar y ajustar reduciendo la

varianza del modelo entrenado, ya que se basa en la sabiduría de las multitudes. La baja correlación entre modelos es la clave. Los modelos no correlacionados pueden producir predicciones de conjunto que son más precisas que cualquiera de las predicciones individuales (Piñeiro & Nakab, 2022).

Otros algoritmos usados en tareas de clasificación son los métodos Hoisting que buscan que los modelos aprendan de los errores del anterior; Uno de ellos es el Gradient Boosting que es un modelo formado por un conjunto de árboles de decisión individuales, entrenados de forma secuencial, de forma que cada nuevo árbol trata de mejorar los errores de los árboles anteriores. La predicción de una nueva observación se obtiene combinando las predicciones de todos los árboles individuales que forman el modelo (Piñeiro & Nakab, 2022).

3 Metodología

Para el desarrollo del proyecto se plantearon 4 etapas de trabajo con el fin de abarcar adecuadamente todos los pasos para realizar el modelo analítico del score de medición de riesgo LAFT en las jurisdicciones nacionales.

Etapas 1. Indagación de la información: En la etapa 1 se realizó todo el análisis y selección de variables aplicables al modelo, estas variables fueron definidas a partir de los delitos fuente del lavado de activos del artículo 323 del código penal. Acá se determinaron qué aspectos específicos se deben tener en cuenta para evaluar el riesgo LAFT en cada jurisdicción. Además, se realizó la recolección de la información de las bases de datos de los entes gubernamentales y con fecha de actualización 2020 y 2022.

Etapas 2. Modelado: Después de tener las variables para el modelo seleccionadas, y la información recolectada con esta información se crearon dos bases de datos una base de datos para el año 2020 y otra base de datos con los datos del 2022, luego de haber construido las bases de datos, se realizó el preprocesamiento y un pequeño análisis exploratorio.

Cuando ya se tuvieron los datos organizados se usaron herramientas de Machine Learning con algoritmos de aprendizaje supervisado: Decision Tree Classifier, Random Forest, Gradient Boostier Classifier, con el fin de que estos nos ayudaran a la clasificación las jurisdicciones en alto, medio o bajo según sea el caso con la mejor precisión posible.

Etapas 3. Validación Para esta validación se usaron métricas de desempeño de los modelos de clasificación como el accuracy que es la métrica que representa el valor el porcentaje total de valores correctamente clasificados, la precisión que es usada para saber el porcentaje de valores

que se han clasificado como positivos y realmente son positivos, el recall que es usada para saber cuántos valores positivos son correctamente clasificados, el F1 score que es una métrica que combina la precisión y el recall, además se usaron métricas gráficas como la matriz de confusión y la curva ROC y con esta el AUC que es el área bajo la curva, estas también usadas para evaluar el rendimiento de los modelos en cuanto a valores clasificados correctamente.

Después de seleccionar el modelo con mejor desempeño, se realizaron validaciones a la clasificación ya que por conocimiento en el negocio se conocen algunos municipios muy riesgosos en LAFT que se esperaban estuvieran clasificados como riesgo algo esto con el fin de ver la confiabilidad en el modelo ya que esto es esencial para tomar decisiones informadas y efectivas en la gestión del riesgo, ya que se busca identificar de manera precisa las jurisdicciones y áreas de mayor riesgo.

Etapa 4. Divulgación Luego de tener la solución analítica con los municipios clasificados se realizó una herramienta de usuario final con la que se realizó divulgación de los resultados y se dio instrucciones a los usuarios de cómo se utiliza.

Además, se realizó la correcta documentación de todo el proceso que se llevó a cabo para la realización del proyecto.

4 Resultados y análisis

Partiendo de la metodología y dando cumplimiento al orden establecido se obtienen los siguientes resultados de cada etapa:

4.1 Indagación de la información

Para realizar la indagación de información se partió del código penal en el artículo 323 (Ley 599 del 2000) del cual se obtienen los delitos fuentes de lavado de activos:

- Trafico de migrantes
- Trata de personas
- Extorsión
- Enriquecimiento ilícito
- Secuestro extorsivo
- Rebelión
- Tráfico de armas
- Financiación del terrorismo y administración de recursos relacionados con actividades terroristas
- Delitos contra el sistema financiero
- Delitos contra la administración fuente
- Delitos producto de un concierto para delinquir

A partir de esta lista de delitos se realizó una lluvia de ideas de las variables que se podrían usar para el modelo. Luego de la lluvia de ideas se tuvieron alrededor de 43 variables posibles, pero que finalmente se fueron descartando por la falta de datos en las bases de datos publicas

DESARROLLO DE MODELO DE MEDICIÓN DE RIESGO DE LAVADO DE ACTIVOS Y FINANCIACIÓN DEL TERRORISMO EN LAS JURISDICCIONES NACIONALES.

gubernamentales y/o por la falta de credibilidad de algunas fuentes, también se descartaron variables por no tener información para cada jurisdicción, finalmente las variables seleccionadas son las que encontramos en la Tabla 1 donde se describe cada una de ellas:

Tabla 1

Variables Seleccionadas

Nombre	Descripción
Cultivos de Cocaína.	Hectáreas de cultivos de cocaína en cada jurisdicción.
Marihuana Incautada	Kilogramos de marihuana incautada en cada jurisdicción.
Base de coca incautada	Kilogramos de Cocaína incautada en cada jurisdicción.
Otras Drogas incautadas	Kilogramos de otros tipos de drogas (Basuco, éxtasis, heroína) incautados en cada jurisdicción.
Laboratorios para la producción de drogas desmantelados	Cantidad de laboratorios para la producción de drogas que son desmantelados en cada jurisdicción.
Armamento incautado	Cantidad de armas y/o objetos usados por grupos armados incautados en cada jurisdicción en los años.
Municipio con grupos armados	Esta variable categórica nos dice si en la jurisdicción el conflicto armado esta: <ul style="list-style-type: none"> •Conflicto activo •Sin conflicto
Riesgo electoral	Variable categórica que mide el riesgo electoral de cada jurisdicción para cámara y senado: <ul style="list-style-type: none"> •Riesgo alto •Riesgo medio •Riesgo bajo
Índice riesgo victimización	El Índice de Riesgo de Victimización (IRV) es una herramienta para el análisis del riesgo de violaciones a los Derechos Humanos e infracciones al Derecho Internacional Humanitario a nivel municipal. <ul style="list-style-type: none"> •Riesgo alto •Riesgo medio-alto •Riesgo medio •Riesgo medio-bajo •Riesgo bajo
Extorsión	Cantidad de extorsiones en cada jurisdicción.
Capturas relacionadas con delitos del código penal el artículo 323	Cantidad de personas capturadas por delitos relacionados al LAFT: <ul style="list-style-type: none"> •Lavado de activos •Tráfico de migrantes y trata de personas •Tráfico de moneda •Financiación del terrorismo •Fabricación y tráfico de armas

**DESARROLLO DE MODELO DE MEDICIÓN DE RIESGO DE LAVADO DE ACTIVOS
Y FINANCIACIÓN DEL TERRORISMO EN LAS JURISDICCIONES NACIONALES.**

	<ul style="list-style-type: none"> •Concierto para delinquir •Rebelión
Capturas relacionadas a delitos ambientales	Cantidad de personas capturadas por delitos medio ambientales relacionados a LAFT: <ul style="list-style-type: none"> •Deforestación •Explotación ilegal de yacimientos de minerales precioso •Contaminación ambiental por explotación de recursos naturales para explotación ilegal de minerales preciosos.
Municipio fronterizo	Zona fronteriza es toda la que se halla cercana a los límites, donde existe entre los habitantes de esas zonas limítrofes en un gran contacto de tipo cultural, económico y comercial.
Municipios con puertos	Indica si el municipio tiene puestos marítimos.
Municipio con presencia ríos	Indica si el municipio tiene ríos principales para realizar desplazamiento fluvial.
Municipio con zona franca	Indica si el municipio tiene Zonas francas, estas son áreas geográficas delimitadas del territorio nacional, en donde se desarrollan actividades industriales de bienes y de servicios o actividades comerciales, bajo una normatividad especial en materia tributaria, aduanera y de comercio exterior.
Municipio con minero	Indica si el municipio tiene como actividad comercial la extracción de minerales.
Municipio con deforestación	Municipios que hacen parte de los núcleos de deforestación.
ROS (Reporte de operaciones sospechosas)	Cantidad de reportes de operaciones sospechosas realizadas en el municipio
Giros Enviados	Cantidad de giros que se enviaron en el año desde ese municipio
Giros Recibidos	Cantidad de giros que se recibieron en el año en ese municipio
Remesas recibidas	Cantidad de remesas que se recibieron en el año en ese municipio
Clientes con actividades de alto riesgo	Clientes con actividades catalogadas por la organización como actividades de alto riesgo.

Nota: Elaboración propia

Las fuentes de los datos usadas fueron:

- Policía Nacional de Colombia
- Ministerios de Colombia (GOV.CO)
- Misión de observación electoral (MOE)
- Centro de análisis de conflictos (CERAC)
- Departamento administrativo nacional de estadísticas (DANE)

Para la limpieza se verifica que los datos estén organizados y que estén completos para cada una de las jurisdicciones es decir no haya datos nulos, en el caso de que en la jurisdicción no se presente información en alguna variable se coloca cero.

Para el preprocesamiento de los datos se realiza una función para convertir los datos en datos per cápita con el fin de que puedan tener una escalabilidad por los diferentes tamaños de las poblaciones para esto se usó la siguiente formula:

$$Per_{capita} = \frac{Cantidad\ de\ eventos}{Población\ del\ municipio}$$

Para el análisis exploratorio se realiza análisis a las distribuciones de las variables como se puede observar en la Figura 2; donde se observa que la mayoría de ellas tienen un sesgo a la derecha, lo que nos indica que la cola derecha de la distribución es más larga o pesada que la cola izquierda. Esto es debido a que hay valores extremadamente altos (outliers) que están estirando la distribución en la dirección positiva esto se da debido a que hay municipios donde los valores son muy altos y otros donde son demasiado bajos o incluso cero.

Figura 2.

Distribución de las variables.

DESARROLLO DE MODELO DE MEDICIÓN DE RIESGO DE LAVADO DE ACTIVOS Y FINANCIACIÓN DEL TERRORISMO EN LAS JURISDICCIONES NACIONALES.



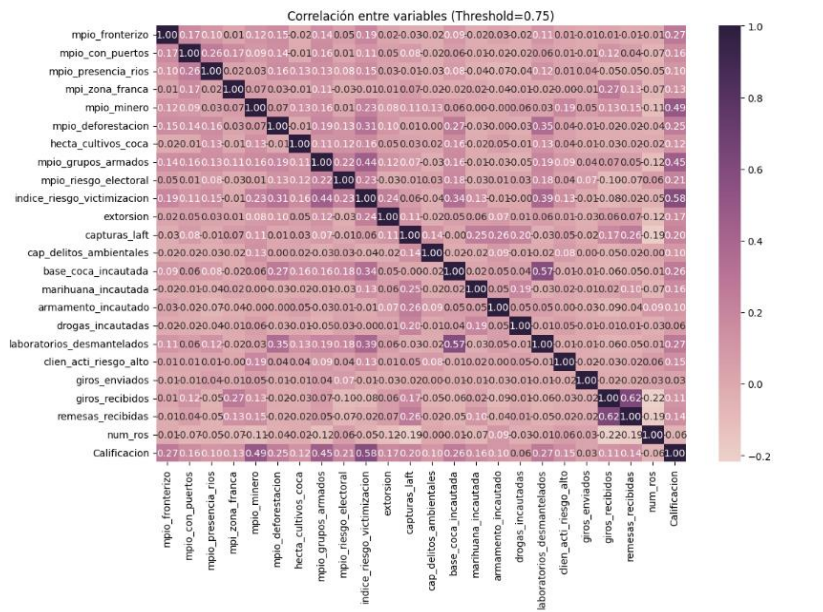
Nota Fuente: Elaboración propia

Se realizó diagrama de correlaciones Figura 3 donde se puede observar que las variables están poco correlacionadas entre ellas, lo que nos indica que no hay una relación lineal fuerte entre las variables, esto implica que los cambios en una variable no están sistemáticamente relacionados con los cambios en otra variable, lo que es muy importante para el modelado ya que en caso de estar muy correlacionadas se agregaría una complejidad extra al modelo ya que tendríamos problemas de multicolinealidad, lo que dificultaría la interpretabilidad del modelo, la precisión en las predicciones ya que una variable estaría explicando en gran medida el modelo lo que lo haría volverse menos robusto.

Figura 3.

Diagrama de correlación.

DESARROLLO DE MODELO DE MEDICIÓN DE RIESGO DE LAVADO DE ACTIVOS Y FINANCIACIÓN DEL TERRORISMO EN LAS JURISDICCIONES NACIONALES.



Nota Fuente: Elaboración propia.

4.2 Modelado

Como el objetivo principal del modelo es definir para cada uno de los municipios de Colombia si tienen riesgo alto, medio o bajo. Teniendo en cuenta esto, se usaron herramientas de Machine Learning como lo son el Aprendizaje supervisado siendo los algoritmos usados el Decision Tree Classifier, el Random Forest y el Gradient Boostier Classifier.

Para realizar el modelado primero se creó la variable de respuesta con la clasificación inicial, cuando ya se tuvo la variable de respuesta se realizó el entrenamiento del modelo con datos del 2020 y finalmente se realizó la predicción en el conjunto de datos con fecha 2022, con el fin de obtener la clasificación para ese año.

Para la creación de la variable de respuesta inicial se usaron los percentiles de las variables numéricas y las coincidencias de los municipios que se encontraban en esos percentiles, esto se llevo a cabo de la siguiente manera:

1. A cada variable numérica se le buscaron los percentiles.
2. Se analizaron los municipios que estaban en esos percentiles para cada variable.
3. Se analizo cuales municipios se repetían por variable en el percentil 90.
4. Los municipios que coincidían en 4 o más variables en el percentil 90 se consideraban municipios de alto riesgo.
5. Los municipios que coincidían de 2 a 4 variables en el percentil 90 se consideraban municipios de riesgo medio.
6. Los municipios que no coincidían en ninguna variable se consideran municipios de riesgo bajo.

Esta metodología dio como resultado la siguiente clasificación inicial:

Municipios ALTO 287

Municipios MEDIO 552

Municipios en BAJO 733

Teniendo la clasificación inicial se procedió al modelado de los datos usando los algoritmos: Decision Tree Classifier, el Random Forest y el Gradient Boostier Classifier en los cuales se realizó entrenamiento con los datos del 2020 y predicción en los datos del 2022 con cada uno de ellos para finalmente comparar cual es el que mejor predicción nos aportó.

Para el proceso de selección del modelo se tuvieron en cuenta las métricas de evaluación de desempeño de las modelos y métricas graficas.

El modelo seleccionado fue el **Gradient Boostier Classifier** con un accuracy del 77% como lo podemos ver en la Figura 4 donde se encuentra el reporte de métricas tanto para el entrenamiento como para la predicción, donde se muestra además la precisión, recall y F1 score, métricas que se tuvieron en cuenta para la selección del modelo.

Figura 4.

Reporte de métricas Gradient Boostier Classifier

```
Test - Accuracy : 0.7733333333333333
Test - classification report :

              0          0.86          0.93          0.90
              1          0.50          0.29          0.37
              2          0.57          0.66          0.61

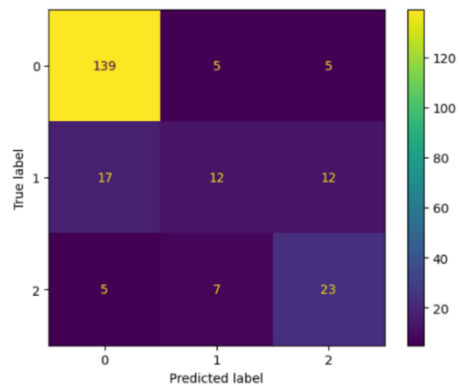
      accuracy                    0.77
      macro avg          0.65          0.63          0.63
      weighted avg       0.75          0.77          0.76
```

Nota Fuente: Elaboración Propia.

Además de las métricas de evaluación también se evaluaron métricas graficas como lo son la matriz de confusión y la curva ROC, figuras 6 y 7 con estas se ayudó a evaluar el rendimiento de los modelos en cuanto a valores clasificados correctamente y los falsos positivos y falsos negativos, la matriz de confusión al ser más intuitivas y sencilla ayudo a interpretar la precisión y exactitud de manera visual y cómo podemos ver en el modelo los falsos positivos y falsos negativos son realmente pocos Figura 5.

Figura 5.

Matriz de Confusión Gradient Boostier Classifier

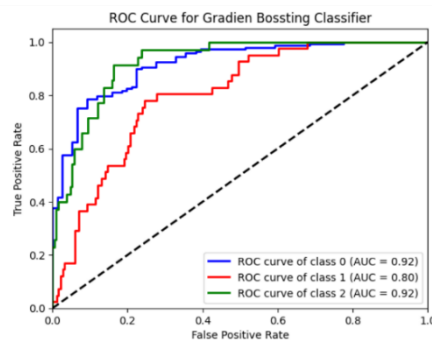


Nota Fuente: Elaboración Propia.

Con la curva ROC (Figura 6) buscamos interpretar que tan bien el modelo puede clasificar una clase de otra, en este caso el tipo de riesgo y el AUC es un puntaje que nos dice que tan bien funciona el modelo, en este caso tenemos una muy buena clasificación para las 3 clases con una curva ROC aproximándose a valores cercanos al 1 que es el ideal y un AUC para las 3 clases sobre el 80%, esto indicando el buen rendimiento del modelo para distinguir una clase de otra.

Figura 6.

Curva ROC Gradient Boostier Classifier



Nota Fuente: Elaboración Propia.

Para la selección del modelo fue parte fundamental el conocimiento de los expertos pues ya se tiene conocimiento de municipios que son riesgoso frente al LAFT en la organización, por lo que en mesa de trabajo con los expertos se validó cuáles eran esos municipios más riesgosos y la clasificación que le dio el modelo y se buscó que el modelo elegido tuviera la clasificación más acertada frente a esto.

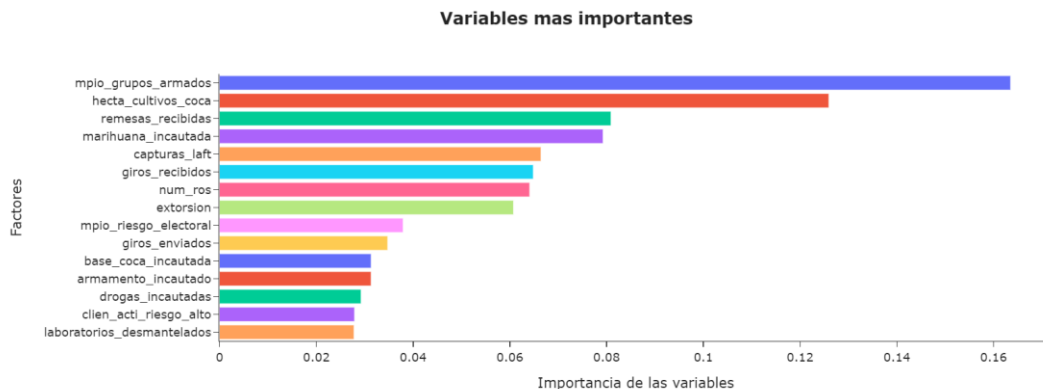
El Gradient Boostier Classifier nos entregó la siguiente clasificación para los municipios:

- Municipios riesgo alto: 205
- Municipios riesgo medio: 179
- Municipios riesgo bajo: 737

Además de la clasificación buscamos resolver una pregunta clave para el negocio ¿Por qué una jurisdicción está en esa clasificación? Esta interrogante buscamos responderla partiendo de las variables y su importancia en el modelo, para esto se categorizaron así: criminalidad, narcotráfico, transaccionalidad interna, contrabando y medio ambiente, con el fin de enfocar su relación con el lavado de activos y la financiación del terrorismo.

Figura 7.

Variables Importantes Modelo Gradient Boostier Classifier



Nota Fuente: Elaboración Propia

DESARROLLO DE MODELO DE MEDICIÓN DE RIESGO DE LAVADO DE ACTIVOS Y FINANCIACIÓN DEL TERRORISMO EN LAS JURISDICCIONES NACIONALES.

En la Figura 7 podemos observar las variables más importantes para el modelo, siendo los grupos armados, las hectáreas de cultivos ilícitos y las remesas las más importantes algo que se esperaba del modelo ya que esas son variables que frente al negocio se conocen como fuente de LAFT En la Tabla 2 podemos encontrar todas las variables por categoría y el peso de importancia.

Tabla 2.

Importancia de las variables por categoría.

Categoría	Variable	Peso
Contrabando	mpio_fronterizo	0.002537
	mpio_con_puertos	0.000627
	mpio_presencia_rios	0.002161
	mpi_zona_franca	0.000096
	mpio_riesgo_electoral	0.037948
Criminalidad	extorsión	0.060746
	armamento incautado	0.031324
	mpio_grupos_armados	0.163489
	indice_riesgo_victimizacion	0.025072
	capturas_laft	0.066427
Variables internas	clien_acti_riesgo_alto	0.027904
	giros_enviados	0.034744
	giros_recibidos	0.064848
	remesas_recibidas	0.080909
	num_ros	0.064098
Narcotráfico	hecta_cultivos_coca	0.125937
	base_coca_incautada	0.03133
	marihuana_incautada	0.079277
	drogas_incautadas	0.029251
	laboratorios_desmantelados	0.02779
Medio Ambiental	mpio_deforestacion	0.000051
	mpio_minero	0.020446
	cap_delitos_ambientales	0.022987

Con estos pesos de importancia para cada variable se puede conocer la importancia por categoría siendo las mayores las relacionadas a criminalidad, narcotráfico y transaccionalidad internas, estos resultandos siendo muy coherentes con lo que se esperaba del modelo y lo que se conoce frente al país y el negocio.

4.3 Herramienta de usuario final

El score de calificación de municipios se presenta en una herramienta de usuario final construida en un tablero de Power Bi, con el fin de que este sea dinámico para los usuarios.

El tablero se presenta de la siguiente manera:

Figura 8.

Pantalla inicio del tablero PBI



Nota Fuente: Elaboración Propia

En la primera página luego de la portada se podrá seleccionar y buscar el municipio y el departamento del cual hace parte, se encuentra el mapa de Colombia y los colores indican que riesgo tiene el municipio y el número de municipios que hay en total en el país, se podrán filtrar por calificación de riesgo o simplemente buscar el municipio que necesite.

Figura 9.

Pantalla del tablero PBI con municipio seleccionado

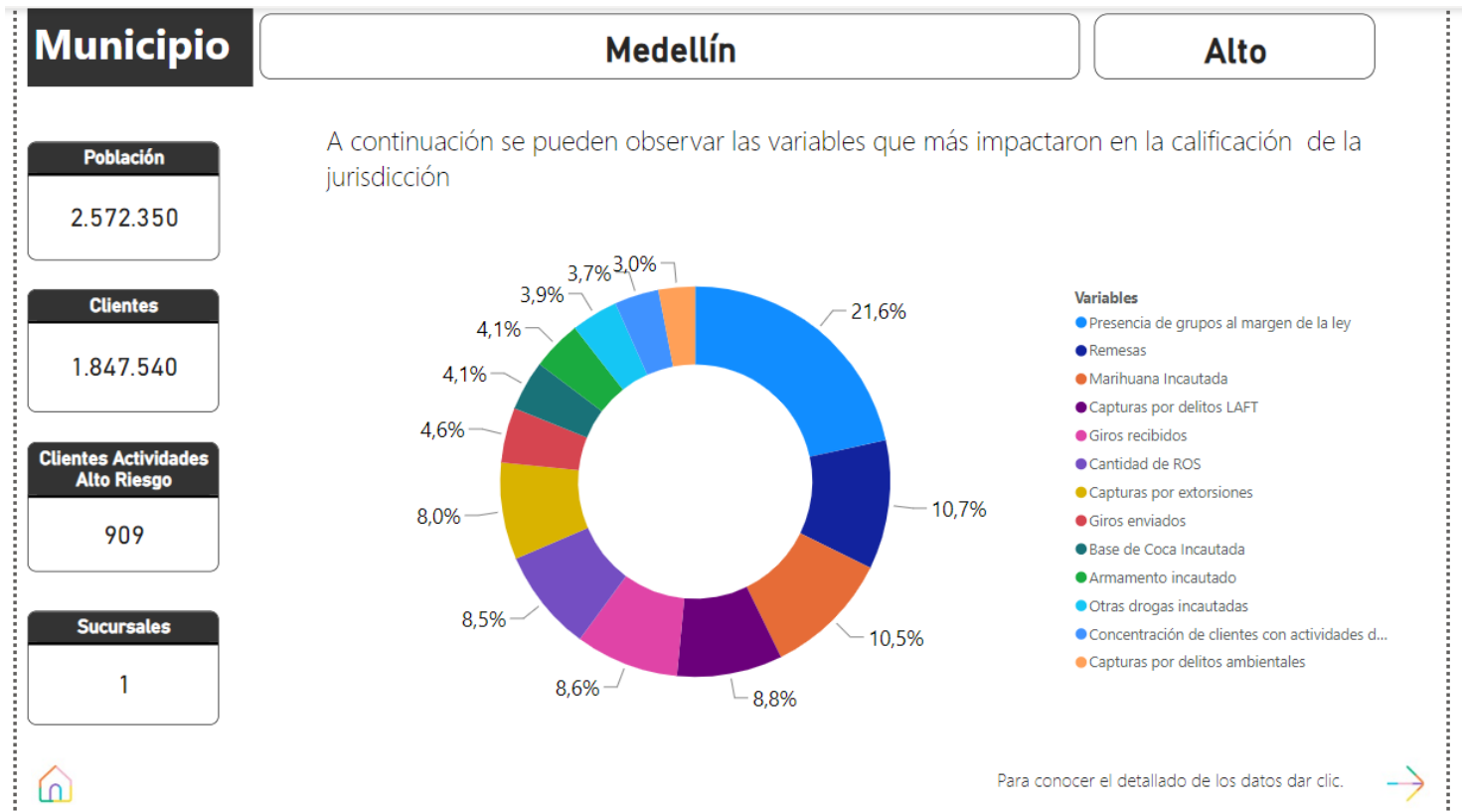


Nota Fuente: Elaboración Propia

Luego de seleccionar un municipio se muestra solo la calificación de ese municipio y se marcara su ubicación en el mapa. Además, para ir a la calificación detallada del municipio se debe dar clic en el botón que ahí se muestra.

Figura 10.

Pantalla calificación del tablero PBI



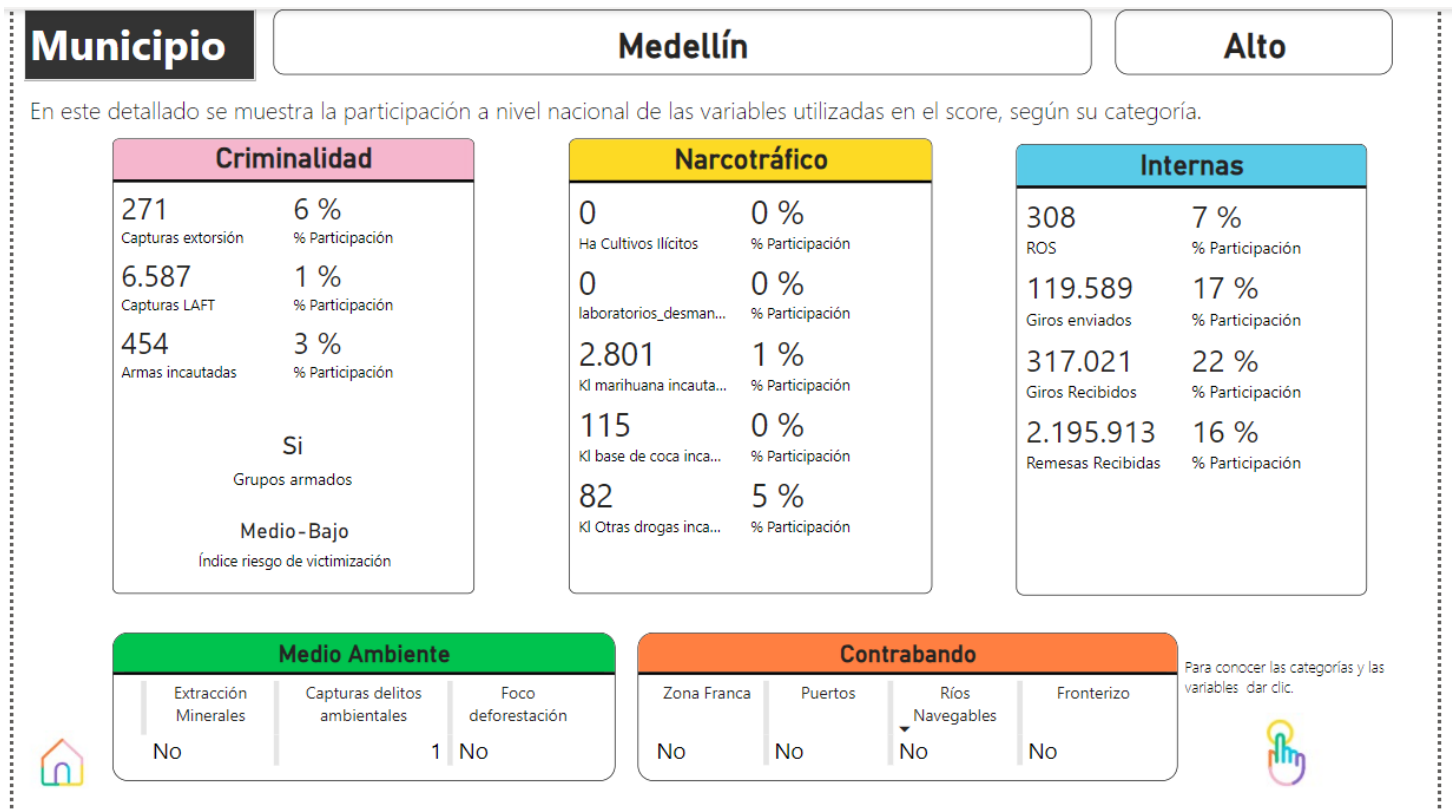
Nota Fuente: Elaboración Propia

Para esta visualización se encuentra información de interés del municipio como su población, la cantidad de clientes que tiene el banco en este municipio, los clientes que tienen actividades de alto riesgo, y el número de sucursales en el municipio, además se muestra cuáles son las variables que más impactaron en la calificación de ese municipio y la manera en la que impactaron.

DESARROLLO DE MODELO DE MEDICIÓN DE RIESGO DE LAVADO DE ACTIVOS Y FINANCIACIÓN DEL TERRORISMO EN LAS JURISDICCIONES NACIONALES.

Figura 11.

Pantalla descarga del tablero PBI



Nota Fuente: Elaboración Propia

En la siguiente visualización se puede encontrar el detallado, acá se ve la información de cada una de las variables por cada categoría y además la participación de cada variable a nivel nacional.

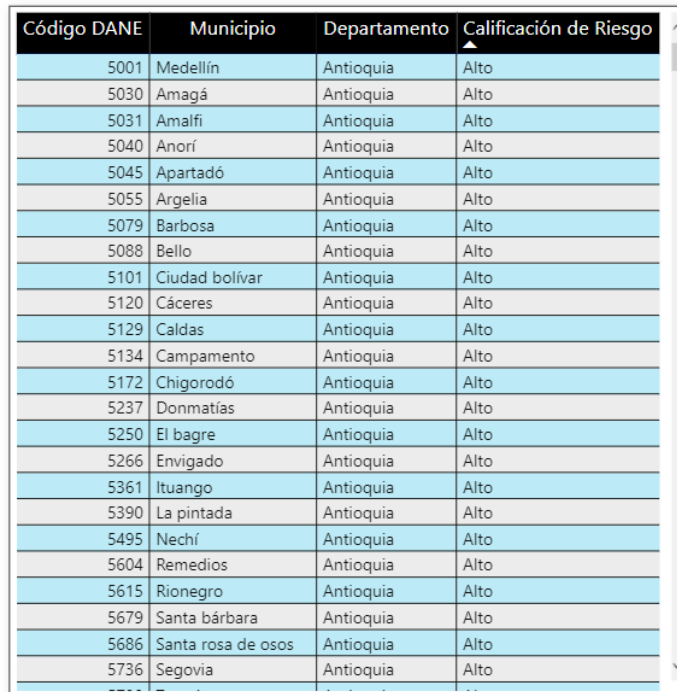
Figura 12.

Pantalla descarga del tablero PBI

Acá podremos descargar el Score de Jurisdicciones Nacionales para el año 2022

Para realizar la descarga masiva del score lo que debes realizar los siguientes pasos:

1. Hacer clic en los tres puntos sobre la tabla.
2. Hacer clic en la opción exportar los datos
3. Guardar el archivo.



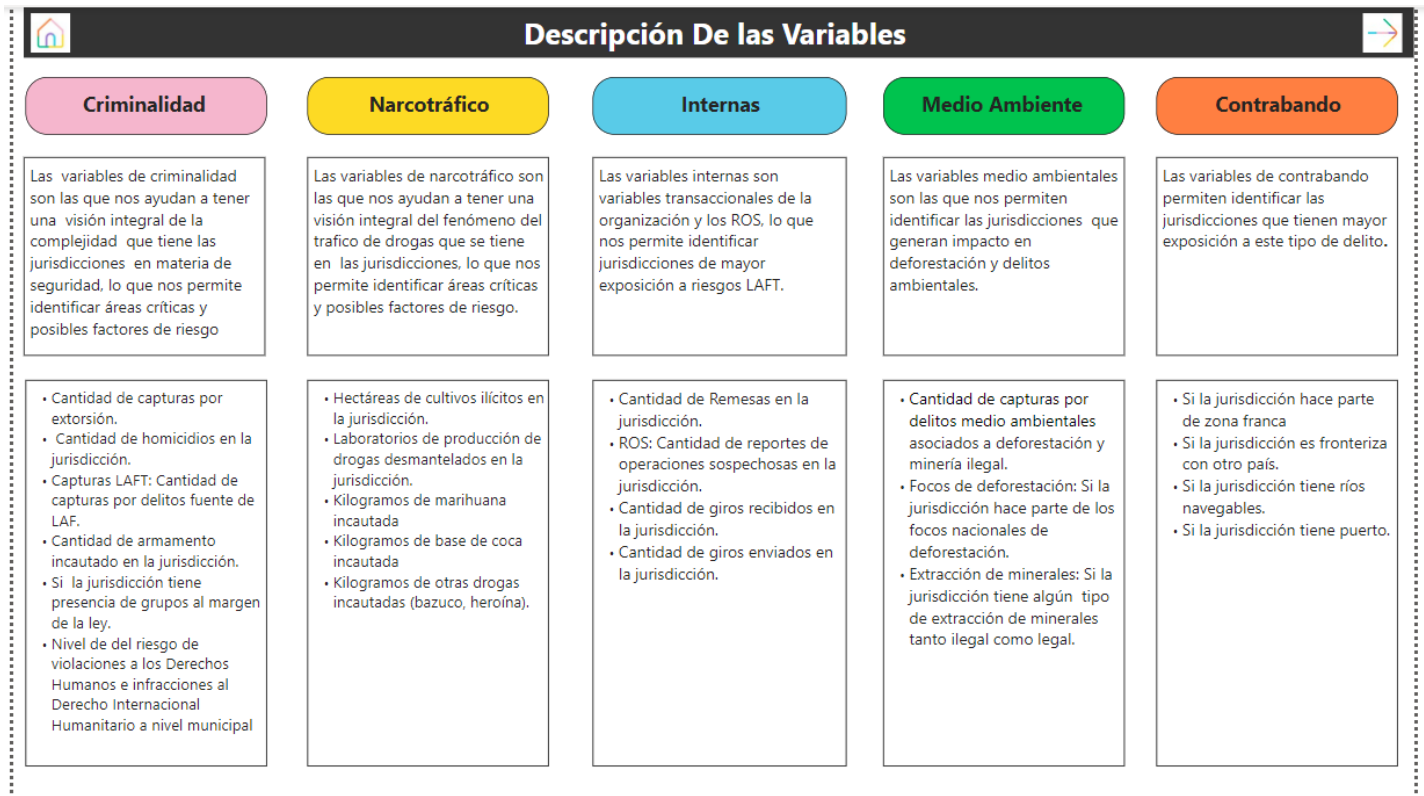
Código DANE	Municipio	Departamento	Calificación de Riesgo
5001	Medellín	Antioquia	Alto
5030	Amagá	Antioquia	Alto
5031	Amalfi	Antioquia	Alto
5040	Anorí	Antioquia	Alto
5045	Apartadó	Antioquia	Alto
5055	Argelia	Antioquia	Alto
5079	Barbosa	Antioquia	Alto
5088	Bello	Antioquia	Alto
5101	Ciudad bolívar	Antioquia	Alto
5120	Cáceres	Antioquia	Alto
5129	Caldas	Antioquia	Alto
5134	Campamento	Antioquia	Alto
5172	Chigorodó	Antioquia	Alto
5237	Donmatías	Antioquia	Alto
5250	El bagre	Antioquia	Alto
5266	Envigado	Antioquia	Alto
5361	Ituango	Antioquia	Alto
5390	La pintada	Antioquia	Alto
5495	Nechí	Antioquia	Alto
5604	Remedios	Antioquia	Alto
5615	Rionegro	Antioquia	Alto
5679	Santa bárbara	Antioquia	Alto
5686	Santa rosa de osos	Antioquia	Alto
5736	Segovia	Antioquia	Alto

Nota Fuente: Elaboración Propia

Se tiene una pantalla de descarga masiva del score pues este score es usado para alimentar otros scores del banco por lo que se necesita una fácil descarga.

Figura 13.

Pantalla descripción de las variables del tablero PBI



Nota Fuente: Elaboración Propia

Finalmente, en la última visualización se puede encontrar una pequeña descripción de las variables para ayudar a mejorar el entendimiento de las variables que usa el score para dar la calificación a la jurisdicción.

5. Conclusiones

En el ciclo de gestión de riesgos, se ha utilizado el manual SARLAFT 4.0 como guía para promover un enfoque basado en riesgo. El Grupo Bancolombia ha demostrado su compromiso al crear políticas, procedimientos, metodologías y herramientas específicas para la identificación, medición, control y monitoreo del riesgo LAFT. La identificación de factores de riesgo clave, como clientes, productos, canales y jurisdicciones, ha sido esencial para una gestión efectiva.

El enfoque particular en las jurisdicciones nacionales, donde el Grupo Bancolombia tiene una presencia significativa, ha llevado al desarrollo de un nuevo score de riesgo LAFT mediante la aplicación de métodos de Machine Learning como herramienta para clasificar las jurisdicciones nacionales.

Para llevar este proceso a cabo se realiza una selección de variables basadas en el artículo 323 del código penal, inicialmente se tuvieron 43 variables y que finalmente las seleccionadas fueron 23, las cuales se categorizaron en: criminalidad, narcotráfico, contrabando, medio ambiente y transaccionalidad interna. Esta selección estratégica proporciono un enfoque más preciso y eficiente para abordar y evaluar aspectos específicos relacionados con dichas categorías en el contexto del lavado de activos y la financiación del terrorismo.

Tras llevar a cabo el proceso de modelado y validación, se determinó que el Gradient Boostier Classifier fue la opción seleccionada. Este modelo identificó 205 municipios con riesgo alto, 179 con riesgo medio y 737 con riesgo bajo. La selección de este modelo se basó en sus métricas de evaluación y rendimiento. Además, la validación mediante una mesa de expertos respaldó la precisión de la clasificación proporcionada por este modelo, reforzando así su precisión para la tarea en cuestión. Este resultado respalda la confianza en la capacidad del Gradient Boostier Classifier para categorizar efectivamente los municipios según su nivel de riesgo.

Finalmente, para la presentación del score se realizó una herramienta de usuario final de fácil uso y acceso y se realizaron las divulgaciones de resultados con las personas de la vicepresidencia.

Este enfoque proporciono al Grupo Bancolombia una herramienta valiosa para tomar decisiones informadas y estratégicas en la prevención y mitigación del riesgo LAFT a nivel nacional, además de no solo contribuir a la seguridad financiera de la organización, sino que

DESARROLLO DE MODELO DE MEDICIÓN DE RIESGO DE LAVADO DE ACTIVOS Y FINANCIACIÓN DEL TERRORISMO EN LAS JURISDICCIONES NACIONALES.

también demuestra su compromiso continuo con el desarrollo sostenible y el bienestar de la sociedad.

Es fundamental destacar que este proyecto no solo ha surgido como respuesta a la necesidad de actualizar la evaluación de riesgos para las jurisdicciones, debido a la falta de actualización del aplicativo proporcionado por el proveedor externo desde 2021, sino que también representa un paso significativo hacia la autonomía y adaptabilidad del Grupo Bancolombia frente a los cambios dinámicos en las jurisdicciones. La capacidad de la organización para generar un nuevo score de jurisdicciones nacionales demuestra un enfoque proactivo en la identificación y gestión de riesgos emergentes permitiendo una respuesta más ágil y efectiva frente a posibles amenazas.

6. Recomendaciones

Dada la relevancia crítica de la información pública sobre criminalidad y narcotráfico para la correcta ejecución de este proyecto, se recomienda enfocar esfuerzos en mejorar la recopilación de datos para cada municipio.

La falta de información para algunos municipios resultó en la exclusión de variables esenciales, lo cual puede afectar la integridad y precisión del modelo.

Algo fundamental para la información que se usa es que sea de fuentes gubernamentales por lo que establecer una colaboración estrecha con entidades gubernamentales y organismos de seguridad para facilitar el acceso a datos cruciales ayudaría a resolver problemas en la recolección de los mismo y se tendría la certeza de la veracidad de los mismos, ya que la mejora en la recopilación de información fortalecerá la base de datos del proyecto, permitiendo una evaluación más precisa y completa de los riesgos LAFT en los municipios.

7.Referencias

- Bancolombia. (2022). *Informe metodológico aplicación del SARLAF en el grupo Bancolombia*. Medellín.
- Cánovas, A. A., Cubo, E., Mir Rivera , P., & Ferro, Á. (2023). *Manual de nuevas tecnologías en transtorno del movimiento* . Madrid : Ediciones SEN .
- Castro, J. A. (2022). *Aplicación de machine learning en la gestión de riesgo de crédito financiero: Una revisión sistemática*. Lima, Peru: Revista Interfases.
- Colombia, B. d. (2014). *Vista de Programa de Revisión del Sistema de Administración del Riesgo de Lavado de Activos y de la Financiación del Terrorismo – SARLAFT*. Bogotá.
- Colombia, B. d. (2014). *Vista de Programa de Revisión del Sistema de Administración del Riesgo de Lavado de Activos y de la Financiación del Terrorismo – SARLAFT*.
- Colombia, S. f. (2022). *Circular Externa 011 de 2022*. Bogotá.
- Colombia, S. F. (2023). *Superintendencia financiera de colombia* .
- Marrugat, A. S. (2020). *Comparación de algoritmos de clasificación* . Barcelona: ETSEIB.
- Mustafa , E. (2019). What is Artificial Intelligence? Technical Considerations and future perception. *Anatol J Cardiol* 2019; 22: 5-.
- Parra, F. (2019). *Estadística y Machine Learning con R*. Cantabria: Ediciones Académicas.
- Piñeiro, S., & Nakab, A. (2022). *Árboles de Decisión: Predicciones como alertas del sistema de prevención de lavado de activos en una entidad financiera*. Obtenido de Universidad Torcuato Di Tella: <https://repositorio.utdt.edu/handle/20.500.13098/11864>

**DESARROLLO DE MODELO DE MEDICIÓN DE RIESGO DE LAVADO DE ACTIVOS
Y FINANCIACIÓN DEL TERRORISMO EN LAS JURISDICCIONES NACIONALES.**

Rodrigo, J. A. (2 de 12 de 2023). *ciencia de datos*. Obtenido de https://cienciadedatos.net/documentos/py09_gradient_boosting_python

Rodrigo, J. A. (3 de 12 de 2023). *Ciencias de datos* . Obtenido de https://cienciadedatos.net/documentos/py08_random_forest_python