

INSTITUCIÓN UNIVERSITARIA POLITÉCNICO

GRANCOLOMBIANO

FACULTAD DE INGENIERÍA, DISEÑO, E INNOVACIÓN

MAESTRÍA EN INGENIERÍA DE SISTEMAS

GRUPO DE INVESTIGACIÓN CITIC

LÍNEA DE PROFUNDIZACIÓN: CIENCIA DE DATOS

**“Aplicación de la teoría de grafos y la inteligencia artificial en el manejo
procesal, como una herramienta de apoyo en los procesos de
responsabilidad fiscal”**

PRESENTA: LUIS ERNESTO BOCANEGRA RAIRÁN

2010020013

ASESOR TEMÁTICO:

JOSÉ MANUEL CHAUTA TORRES. PhD

ENERO DE 2022

Dedicatoria

Primero que todo a Dios y la Virgen, a mi familia, mis padres, esposa e hijas, por su paciencia en el tiempo que pasé en el desarrollo de este proyecto este honor es de ustedes.

AGRADECIMIENTOS

Quiero agradecer muy especialmente a Dios y la Virgen, por la vida, permitirme llegar con salud y ánimo a estas instancias y poner siempre un ángel de aliento cuando creía desfallecer.

A mis queridos y padres Luis Augusto Bocanegra (Q.E.P.D), Alicia Rairán Camacho, al primero por su apoyo económico y sacrificio, a pesar que ya no este conmigo, a la segunda por su amor, esfuerzo y dedicación, este título es tuyo mamita.

A mi esposa Nidia Torres, por su paciencia, apoyo y conocimiento, sabes que eres una gran mujer y en más de 16 años, siempre me has acompañado y ayudado en mis locuras, te amo.

A mis Hijas Laura Alejandra y Mariana Valentina, por su tolerancia y autodisciplina, sin su apoyo mis muñequitas esto no hubiera sido posible.

Al profesor José Manuel Chauta y en general a todo el grupo de profesores del Politécnico Gran Colombiano, porque a pesar de la difícil situación de pandemia, estuvieron prestos para colaborarme en las inquietudes presentadas.

A la universidad Minuto de Dios, por la inspiración de buscar mis sueños y sus 6 años como docente, que me permitieron entender la necesidad de llegar hasta estas instancias para seguir avanzando.

A mi jefe en Cromasoft Ing. German Tabares, por su apoyo y comprensión, en estos casi 8 años de labores con la organización, he encontrado más que un jefe, un amigo incondicional que siempre se ha puesto en mis zapatos y me ha apoyado para salir adelante.

A la Auditoria General de la República y sus funcionarios, por estar prestos a resolver las dudas y solicitudes de información, de igual manera por servir de enlace con las contralorías territoriales e intentar persuadirlas para que logaran contestar las inquietudes que durante la realización del proyecto se fueron presentando.

Finalmente, a la Contraloría de Bogotá y sus funcionarios pues fue allí donde nacieron las ideas de este proyecto.

RESUMEN

La corrupción es un flagelo que afecta los intereses de todos. Las medidas establecidas en el marco legal colombiano con la formulación de leyes, documentos de política, creación de organizaciones y endurecimiento de sanciones para actos relacionados con la corrupción, tienen una escasa efectividad de recuperación del erario público (Isaza E., 2011) . Las entidades encargadas del control fiscal, catalogadas como entidades técnicas (Colombia, 1991), no están dotadas tecnológicamente para predecir de manera oportuna la terminación anormal de los procesos fiscales que tienen a su cargo. En el presente documento se demuestra como la utilización de una solución informática basada en la teoría de grafos, la inteligencia artificial y su aplicación mediante la técnica de aprendizaje automático, en las actuaciones procesales de los expedientes de responsabilidad fiscal en las contralorías territoriales, puede llegar a ser una herramienta de apoyo a la gestión misional del control fiscal en beneficio de la sociedad.

ABSTRACT

Corruption is a scourge that affects the interests of all. The measures established in the Colombian legal framework with the formulation of laws, public politic documents, creation of organizations and tightening of sanctions for acts related to corruption, have little effectiveness in recovering the public treasury. (Isaza E., 2011) The entities in charge of fiscal control, classified as technical entities (Colombia, 1991) are not technologically equipped to predict in a timely manner the abnormal termination of the fiscal processes they are in charge of. This document demonstrates how the use of a computer solution based on

graph theory, artificial intelligence and its application, through the technique of automatic learning in the procedural actions of fiscal responsibility files in the territorial comptrollers' offices, can reach be a tool to support the missionary management of the fiscal control for the benefit of society.

PALABRAS CLAVE

Corrupción Gubernamental, fraude, Crisp DM, Aprendizaje automático, Contralorías

KEY WORDS

Government Corruption, fraud, Crisp DM, Machine learning, Comptrollers

ÍNDICE GENERAL

INTRODUCCIÓN	1
1. JUSTIFICACIÓN.....	2
2. OBJETIVOS.....	7
2.1. Objetivo general	7
2.2. Objetivos específicos.....	7
3. ALCANCE DEL PROYECTO	8
4. JUSTIFICACIÓN DEL PROYECTO.....	9
5. PREGUNTA DE INVESTIGACIÓN	13
6. REVISIÓN DE LITERATURA	14
6.1. Problemática.....	14
6.2 Los entes de control	15
6.3 Detección del fraude	15
6.4 Uso de la información.....	17
6.5 Organizar la información	18
6.6 Grafo	19

6.7	Manejo del proceso de responsabilidad fiscal	21
6.8	Aprendizaje de máquina	30
6.9	Aprendizaje Automático	31
6.10	Algoritmos de Minería de Datos	32
6.11	Grafos de conocimiento	33
6.12	Bases de datos orientadas a grafos	34
6.13	Implementación en tramas de corrupción	35
6.14	Balanceo de datos.....	38
7.	ESTRATEGIA METODOLÓGICA	40
8.	CRISP-DM (Cross Industry Standard Process for Data Mining).....	42
9.	DESARROLLO E IMPLEMENTACIÓN	46
9.1.	Análisis de la información.....	46
9.2	Resultados en el análisis con Grafos.....	54
9.3	Validación de Reglas del negocio y limpieza de datos.....	62
9.4	Casos de Uso.....	69
9.5	Diseño de Pantallas	73

10	RESULTADOS	77
11	DISCUSIÓN Y CONCLUSIONES	86
	REFERENCIAS	89
	ANEXO 1	100
	ANEXO 2	102

ÍNDICE DE TABLAS

Tabla 1 Registros por año Información SIA-SIREL	46
Tabla 2 Número de expedientes con causal de prescripción discriminados por año.....	48
Tabla 3 Procesos de Responsabilidad Fiscal con causal de prescripción por contraloría .	48
Tabla 4 Cuantías de procesos prescritos	50
Tabla 5 Concordancia de actuaciones de apertura con columnas SIA-SIREL.....	58
Tabla 6 Concordancia de actuaciones del archivo con columnas SIA-SIREL.....	59
Tabla 7 Concordancia de actuaciones de la imputación con columnas SIA-SIREL.....	60

ÍNDICE DE FIGURAS

Figura 1 Actuaciones generales del Proceso de Responsabilidad Fiscal.....	5
Figura 2 Actuaciones del Proceso de Responsabilidad Fiscal.....	21
Figura 3 <i>Representación de nodo y arista</i>	22
Figura 4 <i>Grafo de actuaciones ley 610 de 2000</i>	23
Figura 5 Grafo etapa de apertura proceso de responsabilidad fiscal	24
Figura 6 <i>Grafo etapa imputación proceso de responsabilidad fiscal</i>	25
Figura 7 Grafo etapa Fallo con Proceso de Responsabilidad Fiscal	26
Figura 8 Grafo Etapa Archivo Proceso de Responsabilidad Fiscal.....	27
Figura 9 Grafo Etapa Fallo Sin Proceso de Responsabilidad Fiscal	28
Figura 10 <i>Operación de relaciones</i>	36
Figura 11 Estrategia Metodológica	40
Figura 12 Metodología CRIS-DM	42
Figura 13 Procesos terminados por Prescripción Procesos de Responsabilidad Fiscal Registrados en el Aplicativo SIA-SIREL.....	50
Figura 14 Cuantías dejadas de recuperar por la Prescripción Procesos de Responsabilidad Fiscal.....	52

Figura 15 Afectación por año de la prescripción Procesos de Responsabilidad Fiscal.....	52
Figura 16 Grafo con los procesos por prescripción encontrados en SIA-SIREL.....	55
Figura 17 Grafo etapas de apertura procesos con prescripción encontrados en SIA-SIREL.	56
Figura 18 Grafo etapas de apertura, archivo e imputación de procesos con prescripción encontrados en SIA-SIREL.	57
Figura 19 Caso de uso acceder al sistema	69
Figura 20 Caso de uso adjuntar Excel con vigencia a analizar	70
Figura 21 Caso de Uso Analizar la coherencia del archivo.....	72
Figura 22 Pantalla de ingreso	73
Figura 23 Pantalla cargue del archivo	74
Figura 24 Pantalla de Reporte de Predicción	75
Figura 25 Columnas tomadas para los experimentos	77
Figura 26 Grafica de exactitud en el modelo 1.....	78
Figura 27 Matrix de confusión Modelo de regresión logística.	79
Figura 28 Matrix de confusión Modelo de Gradient Boosted Tree.	79
Figura 29 Grafica de precisión con el modelo 1.....	80

Figura 30 Grafica de sensibilidad con el modelo 1	80
Figura 31 <i>Grafica de Exactitud con el modelo 2</i>	81
Figura 32 Grafica de Precisión con el modelo 2	82
Figura 33 Grafica de Sensibilidad con el modelo 2.	82
Figura 34 Grafica de clasificador tomado para el modelo 3	83
Figura 35 Grafica de sensibilidad para el modelo 3	84
Figura 36 Grafica de Exactitud para el modelo 3	84
Figura 37 Grafica de Precisión para el modelo 3	85

INTRODUCCIÓN

En el presente documento se presenta diferentes técnicas de inteligencia artificial, utilizadas en el manejo de expedientes procesales, partiendo de analizar las etapas por las que transita un proceso de responsabilidad fiscal, así como también desde la teoría de grafos, las relaciones que se pueden llegar a presentar y como se relaciona la información entre sus nodos.

Por otra parte, se habla de la relación de la corrupción como flagelo que afecta pueblos enteros, quitándole oportunidades para el desarrollo pleno de la sociedad (Palestina, 2018). De igual manera enumerar las herramientas TIC desarrolladas como apoyo a entidades gubernamentales en la lucha contra la corrupción, para finalizar, desde la óptica de las ciencias de la computación los beneficios de implementar soluciones tecnológicas con el ánimo de ser más eficaces al momento de poder mitigar este fenómeno.

1. JUSTIFICACIÓN

Desde la antigüedad, los ciudadanos del mundo han visto como la corrupción los ha acompañado y evolucionado a través del tiempo, su influencia ha permeado las diversas ramas del poder y ha hecho que la ciudadanía cada día crea menos en las instituciones (Amoedo, 2018). La detección del fraude se plantea como uno de los principales desafíos de la administración pública a nivel global. La ciudadanía exige de sus gobernantes, no solo la persecución del delito, sino la prevención, supervisión y la transparencia de la información generada en los entes encargados de esta labor. Esto con el fin de que la sociedad pueda ejercer la veeduría correspondiente y poder denunciar oportunamente, ante las entidades de control, los hechos que intenten desviar el patrimonio público (Zuleta, 2015) (Solé, 2018).

Organizaciones internacionales como la OCDE, la ONU y la Comisión Europea, han alentado a las naciones a que se preste una mayor atención a desarrollar estrategias a la medida, que permitan a las administraciones regionales y locales disminuir de manera significativa los hechos de corrupción, impactando en términos de carencia de servicios a los ciudadanos que previamente han contribuido con sus impuestos (Cerillo i & Martínez, 2018) (Torres De Durand, 2019).

En Colombia este fenómeno no ha sido ajeno. Como parte de las políticas institucionales y pedagógicas en el año 2011 mediante la ley 1474, se estableció el “Estatuto anticorrupción”. Esta ley, en su artículo 73, enmarca el plan anticorrupción

y de atención al ciudadano por parte de todas las entidades del estado. Esto con el objetivo de enfrentar la corrupción y mejorar la gestión pública (Colombia, 2011).

Tecnológicamente, el estatuto obliga a las entidades del estado a que la información relacionada con el mapa de riesgos de la corrupción y su correspondiente plan de mitigación esté publicada en un medio de fácil acceso para el ciudadano. De igual manera que este último, tenga un canal que le permita la gestión de sus denuncias y conocer el tratamiento dentro cada entidad, so pena que el funcionario público que no se acoja a la regulación, sea implicado en un proceso sancionatorio fiscal por violación u omisión como es promulgado en (MinJD. Decreto, 2020) en su Artículo 80, 81 y se hará acreedor de una multa o sanción descrita en el Artículo 84.

En cuanto a la labor de las contralorías está en sujeción constitucionalmente (Colombia, 1991) a la Contraloría General de la República, así como también los entes de control, departamental, municipal, distrital y según las modificaciones de (MinJD. Decreto, 2020) a la Auditoría General de la República. Dicha gestión de los entes de control, se alimentan de las indagaciones y procesos auditores que conllevan a la apertura de procesos de responsabilidad fiscal. Dichos procesos, son manejados por funcionarios sustanciadores de las contralorías en el país, apoyándose de sistemas de información, para el control histórico de los expedientes. Los profesionales sustanciadores actuarán en la figura de policía judicial manejando los expedientes desde el inicio hasta su correspondiente archivo, según se enuncia en (Colombia, 2000), artículo 10.

La propuesta más actual para el mejoramiento del control fiscal se encuentra en el decreto 403 de 2020 (MinJD. Decreto, 2020) que promulga los lineamientos para la implementación del Sistema Nacional de control fiscal - SINACOF – con el que mediante políticas, normas, métodos, procedimientos, herramientas tecnológicas, entre otras, se pretende la unificación, armonización y estandarización de la tarea de vigilancia y control fiscal, así como también análisis y evaluación enfocadas en la gestión y los resultados.

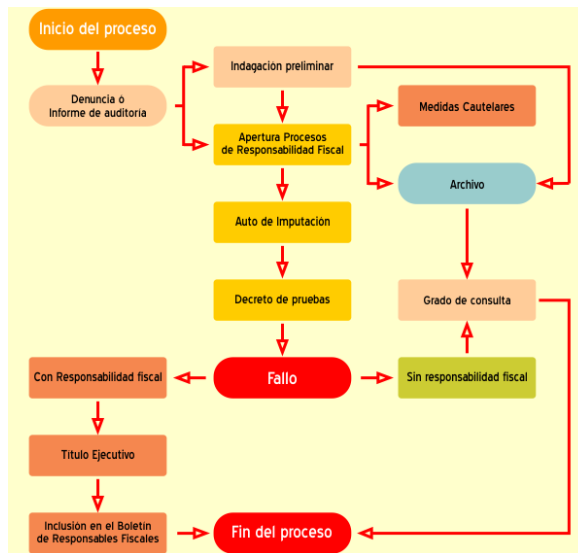
Sin embargo, una de las restricciones que se encuentran en el desarrollo normal de un proceso ordinario es la posible terminación anormal. Entre las terminaciones anormales más relevantes tenemos el pago anticipado de los presuntos responsables fiscales, la caducidad, que hace referencia a un descubrimiento tardío de los hechos que dieron origen al presunto detrimento y la prescripción que según el artículo 9 de (Colombia, 2000) entre la duración de los procesos desde su apertura hasta su decisión de fondo no puede ser superior a 5 años a partir de la apertura del proceso, caso en el cual los investigados pueden solicitar la preclusión de la investigación.

Por otra parte, los procesos de responsabilidad fiscal son vistos desde las entidades de control, quienes se encargan de gestionarlos, como grafos dirigidos, compuestos por un número de actuaciones enmarcadas en la ley 610 de 2000 y que han sido objeto de adiciones y modificaciones a lo largo del tiempo por diferentes leyes, decretos y normatividades internas (CGR, 2021). Un ejemplo de ello lo podemos encontrar en la Contraloría General de la Republica en donde en la siguiente grafica

podemos encontrar como se enuncian las actuaciones más generales de los procesos de responsabilidad fiscal.

Figura 1

Actuaciones generales del Proceso de Responsabilidad Fiscal



Fuente:

<https://cgr.paradigmasln.com/documents/20125/404541/Proceso+de+responsabilidad+Fiscal.png/0f1bec39-701b-8e91-0006-d5ae9b7d81f3?t=1628548238053>

Dados estos antecedentes y en búsqueda de una herramienta tecnológica que mitigue la pérdida fiscal de estos procesos por malas prácticas o corrupción, así como también, que los lineamientos de la implementación del SINACOF, contemplan la integración de la información de muchos sistemas ya implementados en las contralorías territoriales y que aportan a la consecución de resultados en mejora de la gestión fiscal (MinJD. Decreto, 2020), se ha desarrollado como proyecto de investigación la aplicación de la teoría de grafos a la actuación procesal, con el fin de que se genere una respuesta, que mediante el aprendizaje supervisado, permita predecir y detectar de manera oportuna comportamientos procesales, reglamentados en la legislación

nacional tienen un curso normal, pero que desde la óptica analítica contienen determinados caminos de similitud de procesos que han terminado con pérdida fiscal.

Como modelo de implementación se tomarán los datos de procesos de las contralorías del país que tengan como característica propia una terminación anormal, los procesos que tienen un tránsito ajustado en tiempos y los procesos que se encuentran en curso. En la recolección de información, se transforma estadísticamente la misma, limpiando los datos y transformando en una matriz de adyacencia la información para ser graficada en Python mediante el uso de la librería NetworkX, mediante la teoría de grafos se utiliza un algoritmo que analiza la información procesal existente en las bases de datos que manejan las actuaciones procesales de los expedientes fiscales de una contraloría territorial. Posterior a esta recolección y depuración, se procede a hacer una selección de información que alimente un sistema que permita realizar la valoración de riesgos de los procesos en curso y de esta manera determinar una probabilidad de terminación anormal, basándose en el historial de los procesos con terminaciones anormales y las reglas de conocimiento suministradas por la experiencia de los funcionarios sustanciadores. Finalmente, mediante un prototipo de software se mostrarán los resultados utilizando herramientas propias del Big Data para tal fin. Esto llevaría a tratar el problema con el uso de una herramienta de innovación para ser usada por parte de las contralorías territoriales, que ayude a mitigar el fenómeno de la prescripción y mejore la gestión de los entes de control.

2. OBJETIVOS

2.1. Objetivo general

Desarrollar un prototipo de sistema de vigilancia mediante el uso de las ciencias de datos, la Inteligencia Artificial, la teoría de grafos, entre otras técnicas, permita establecer los patrones de comportamiento de los procesos de responsabilidad fiscal en curso, gestionados en las contralorías territoriales del país, contribuyendo a mitigar los casos de terminación anormal.

2.2. Objetivos específicos

1. Analizar las etapas procesales y los factores que podrían generar la terminación anormal de los procesos de responsabilidad fiscal en una contraloría territorial.

2. Extraer información histórica de las bases de datos existentes para determinar los patrones de comportamiento de los procesos con terminación por prescripción.

3. Estimar el nivel de riesgo de terminación anormal de los procesos en curso, mediante un algoritmo basado en la teoría de grafos que utilice los patrones de comportamiento de los procesos con prescripción.

4. Establecer el nivel de precisión del algoritmo mediante la realización de pruebas sobre procesos de responsabilidad fiscal que estén en diferente nivel de riesgo de terminación anormal.

3. ALCANCE DEL PROYECTO

Se pretende la aplicación de las ciencias de datos en el desarrollo de un sistema que permita la vigilancia procesal de las investigaciones en curso en una contraloría territorial. Esto se llevará a cabo mediante la implementación de un algoritmo basado en la teoría de grafos, que analice la información histórica existente y que permita mediante un aprendizaje supervisado, obtener los patrones de comportamiento de los procesos de responsabilidad fiscal que terminan de manera anormal. Con este modelo se procederá a analizar el universo de los procesos en curso, para establecer el nivel de riesgo y contribuir a la toma de decisiones. Para dar cumplimiento a ello se toma la información existente en la herramienta de la Auditoría General de la República SIA-SIREL, que contiene información de 67 contralorías del país.

4. JUSTIFICACIÓN DEL PROYECTO

	¿CÓMO?	¿POR QUÉ?	¿PARA QUÉ?
TÉCNICA	Con la realización de diferentes entrevistas a los actores involucrados en el proceso de responsabilidad fiscal y con visitas técnicas a las direcciones que manejan dicho proceso.	Se hace necesaria la realización de un trabajo de campo con dichos actores para establecer así el manejo dado a la información recibida y qué procesos son afectados con este manejo.	Establecer las tareas y rutinas aplicadas sobre dicha información y que son de conocimiento por parte de la ciudadanía, entes de control, implicados, aseguradoras, entre otros y que mediante mecanismos de participación activa pueden llegar a tener conocimiento.
TECNOLÓGICO	Con la implementación de un algoritmo de búsqueda y levantamiento de información histórica de las bases sistematizadas de una contraloría territorial.	Permitirán rastrear y extraer los procesos con terminaciones anormales, que se encuentren en etapas posteriores para determinar similitudes con los procesos en curso.	Identificar patrones de comportamientos que permitan establecer en los procesos activos posibles riesgos similares de terminación por prescripción.

	¿CÓMO?	¿POR QUÉ?	¿PARA QUÉ?
METODOLÓGICA	Con la utilización de la metodología Crisp DM, que permita el desarrollo de los pasos adecuados para la recolección, preparación de datos e implementación de una solución práctica que apoye la toma de decisiones	Porque esta metodología es la guía de referencia probada en el desarrollo de proyectos de minería de datos y puede ser plenamente adaptada al desarrollo práctico del proyecto.	Para orientar el desarrollo de la propuesta a un esquema probado de éxito y que permita una culminación adecuada del proyecto.
ECONÓMICO	Las herramientas a utilizar en la implementación del proyecto estarán basadas en el desarrollo de software libre.	Mejorar la calidad del trabajo al permitir una mayor eficiencia y eficacia en la toma de decisiones por parte de la alta gerencia del ente de control.	Aportar a la lucha contra la corrupción y a la gestión del ente de control, mejorando la eficacia y el acceso a la información de las actuaciones procesales.

	¿CÓMO?	¿POR QUÉ?	¿PARA QUÉ?
SOCIAL	<p>Aportando en la vigilancia de expedientes, mitigando las falencias presentadas en procesos con terminaciones anormales y alertando en oportunidad malas prácticas que faciliten la terminación anormal.</p>	<p>La corrupción como fenómeno mundial ataca los recursos fiscales de todos los contribuyentes y quita rubros presupuestales en diferentes áreas sociales.</p>	<p>Para optimizar la gestión fiscal y fortalecer los instrumentos de control que permita el manejo transparente de los recursos públicos de la ciudad.</p>
PRÁCTICO	<p>Con la aplicación de los conocimientos adquiridos en la maestría y en la experiencia profesional.</p>	<p>Se requiere aportar desde la academia con la sugerencia de herramientas que optimicen la labor de los entes de control y permitan que la labor de los funcionarios sustanciadores de procesos relacionados con el fisco sea orientada a mejores resultados.</p>	<p>Innovar la forma en que se vigila el tránsito de los expedientes en curso, basados en la historia de los procesos ya culminados, mejorando la forma en que son manejados, aportando en la lucha contra la corrupción.</p>

	¿CÓMO?	¿POR QUÉ?	¿PARA QUÉ?
LEGAL	<p>Teniendo en cuenta:</p> <p>*Art. 61 de la Constitución Nacional de Colombia.</p> <p>La ley 1712 de 2014 Ley de transparencia</p> <p>*Ley 610 de 2000 trámite de los procesos de responsabilidad fiscal de competencia de las contralorías.</p> <p>* Decreto Ley 403 de 2020.</p> <p>Por el cual se dictan normas para la correcta implementación del Acto Legislativo 04 de 2019 y el fortalecimiento del control fiscal.</p> <p>* Ley 44 de 1993, Ley 23 de 1982, Decreto 126 de 1996, Decreto número 1360 de 1989 (Sobre derechos de autor).</p>	<p>Se debe cumplir con la normatividad que rige los procesos de responsabilidad fiscal, la constitución y el debido proceso.</p> <p>Para no incumplir con la normatividad legal colombiana con respecto a derechos de autor</p> <p>Decreto número 1360 de 1989 (Sobre derechos de autor).</p> <p>Conservar los derechos procesales de los presuntos implicados y terceros responsables.</p>	<p>Exista un soporte legal que permita que la reserva del sumario a la que todo presunto implicado tiene derecho.</p> <p>Se cumpla la normatividad vigente sobre el proceso y la investigación desarrollada aporte de normativamente a la lucha contra la corrupción.</p>

5. PREGUNTA DE INVESTIGACIÓN

¿Cómo aplicar la teoría de grafos y la ciencia de datos para apoyar la gestión de los procesos de responsabilidad fiscal, mejorar la toma de decisiones por parte de las contralorías territoriales y aportar a la lucha contra la corrupción?

6. REVISIÓN DE LITERATURA

6.1. Problemática.

La corrupción es un problema mundial que afecta significativamente a los gobiernos y trae consigo un coste elevado para la sociedad y unos impactos ambientales y sociales evidentes [21]. Los ciudadanos ya no creen en las instituciones, 3 de cada 4 ciudadanos de Latinoamérica tienen poca confianza en sus gobiernos y un 80% de los ciudadanos piensan que la corrupción se encuentra en las instituciones públicas. Se generaliza la administración como un nido de corrupción en donde la politiquería y el clientelismo son los únicos beneficiados (CEPAL, 2020).

La ineficiencia de las actividades desarrolladas a nivel institucional, el grado de monopolio de los servicios que prestan, los ingresos que pueden obtenerse en las actividades corruptas, así como la baja probabilidad de ser descubierto, conllevan a que la sociedad comience a convivir con la corrupción y los sistemas creados para la denuncia ciudadana sean menos consultados (Emerio Villamil, 2010).

Esto nos lleva a indicar que a pesar de que la corrupción es un problema a nivel mundial necesita ser tomado muy en cuenta, debido a que los daños hechos por los corruptos, le quitan las oportunidades de crecimiento y desarrollo a la sociedad (Blanes, 2017).

6.2 Los entes de control

Entre las instituciones que se encargan del control fiscal en el país, tenemos la Contraloría General de la República, Las contralorías departamentales, municipales y distritales. Todas ellas tienen están reguladas por la ley y sobre ellas ejerce control la Auditoría General de la República.

La institucionalidad cada día se ve golpeada por los escándalos propios de las mismas en lo que respecta a su contratación y gestión, los entes de control territorial no son ajenos a este fenómeno. Casos como los de Bogotá en el año 2010, en el que el contralor de la ciudad se vio involucrado en lo que se denominó como el cartel de la contratación, han puesto en duda la transparencia y objetividad de las instituciones de control (Jaimes, 2020). Funcionarios como Miguel Ángel Morales Russi (Contralor de Bogotá) y Samuel Moreno Rojas (alcalde), son recordados por estar involucrados en estos hechos.

6.3 Detección del fraude

Uno de los mayores desafíos de los gobiernos es la detección del fraude. Más allá de los procesos de auditoría y de inspección y vigilancia, la sociedad exige resultados en la lucha frontal contra el mismo. La disminución de malas prácticas a nivel de la administración pública posiblemente prevenga la pérdida de los recursos destinados para cubrir las necesidades básicas del estado.

Una de las teorías más relevantes del fraude se denomina el “triángulo del fraude”, esta aporta una justificación en términos psicológicos y sociológicos de las

actividades fraudulentas. Se basa en la existencia de la racionalización, la presión y la oportunidad, estos tres factores inducen al individuo a cometer un fraude en un contexto específico (Calabuig, 2018).

Se define como racionalización, como la capacidad de racionalizar el hecho en concordancia con su código de ética personal para justificar consciente e intencionalmente el acto fraudulento. El poder o presión, se basa en el estímulo que se tiene, ya bien sea posición estratégica dentro de la organización, el trabajo bajo presión o la administración de bienes o servicios. Finalmente, la oportunidad, que se refiere a las circunstancias que hacen viable la perpetración del fraude, por ejemplo, la deficiencia o ausencia de controles, baja supervisión o ineficiencia de administración (Moreno, 2012).

Así como los gobiernos tecnifican sus métodos para perseguir el fraude, de la misma manera crecen las estrategias de los defraudadores para cometer el mismo, logrando de esta manera la evasión de la acción fiscal y produciendo el desfaldo público.

Es aquí donde el manejo de la información y el uso de la misma en la toma de decisiones desarrollan un papel imprescindible para lograr una buena administración que mitigue la pérdida de los recursos y disminuya la malversación de fondos estatales. Por ello se busca que, con la implementación de una herramienta tecnológica, mitigar la prescripción de los procesos de responsabilidad fiscal adelantados en los entes de control y de esta manera, disminuir la oportunidad de que los defraudadores de los

recursos estatales para que puedan cometer hechos que lleven a la no recuperación del erario público.

6.4 Uso de la información

Es importante que la administración en la búsqueda de mitigar los riesgos derivados de las estrategias utilizadas por los defraudadores para el desvío de los recursos, adquiera herramientas que le permitan tener la información ordenada, clasificada y le permita hacer análisis de la misma. Entre más completa sea la información que se tenga de personas, empresas y particulares con los que tenga relación, puede hacer análisis más efectivos, con la ayuda de herramientas informáticas que le permitan detectar y/o alertar de las malas prácticas, omisiones y anomalías (Calabuig, 2018).

En esta búsqueda de organización y análisis de la información el uso de tecnologías de la información para detectar los riesgos de corrupción por parte de las administraciones públicas lleva esta actividad al ámbito de la Administración electrónica, entendida como el proceso de incorporación de nuevas tecnologías de la información y la comunicación en la gestión administrativa con el fin de lograr una buena administración (Solé, 2018).

Cuando se diseñan procedimientos para el análisis de datos, resulta muy útil disponer de un conocimiento sobre la estructura de los mismos. Este conocimiento puede obtenerse de forma genérica, o bien a través procedimientos que caractericen dichos datos a partir de una muestra de ellos (en esta categoría se enmarcan los

algoritmos basados en redes neuronales), o también estableciendo un modelo matemático a priori de los datos, su distribución, sus relaciones, etc. (Calabuig, 2018).

Elegir el mejor algoritmo para una tarea específica de tipo analítico puede ser un gran desafío. Pues, aunque se pueden usar diferentes algoritmos para realizar tareas similares, el resultado generado puede ser completamente diferente, y algunos pueden generar más de un tipo de resultado (Microsoft, 2020). Mediante la medición de varios test de entrenamiento se selecciona el algoritmo que mejor porcentaje de precisión nos presente.

6.5 Organizar la información

La organización de los datos como se enunciaba anteriormente es un proceso no menos que complejo por la cantidad de herramientas existentes, es de saber que los procesos de responsabilidad fiscal tienen una particularidad mencionada por la ley 610 de 2000 (Colombia, 2000) y es que todos comienzan por una etapa inicial compuesta por etapas de citación y notificación a todos los presuntos responsables. De allí pasa a una fase de investigación que traza rutas similares o en conjunto (Depende de la dinámica del proceso). Finalmente, un proceso puede tener diversos desenlaces para cada uno de los involucrados; y no todos ellos pasan por las mismas etapas procesales independientemente si el final es el mismo.

Cuando un proceso por mala praxis o por vencimiento de términos tiene un desenlace anormal (Caducidad, prescripción, anulación) los presuntos implicados tendrán una afectación similar.

En la información suministrada por la Auditoría General de la República se encontró campos relacionados con los procesos de responsabilidad fiscal, que a primera impresión no tenían mayor relevancia para la obtención de información, más adelante, con el proceso de minería de datos, se pudo establecer lo contrario, por lo que se procedió a tener en cuenta, para las etapas posteriores del modelo de Machine Learning y la creación del prototipo.

6.6 Grafo

Se menciona la interpretación práctica que se le ha dado al proceso, el cual encontramos que su comportamiento se asimila a un grafo de actividades. Definimos un grafo, como un conjunto (no vacío) de objetos denominados nodos (o vértices) y una línea que los relaciona, dichas uniones se llaman aristas que pueden ser dirigidas o no dirigidas. Es entonces, la teoría de grafos, el estudio de las relaciones matemáticas que se usan para modelar estas relaciones (Urueta, 2021).

Otra definición a considerar es que los grafos son una forma casi ubicua de pensar en escenarios del mundo real, ya que abstraen los elementos y las relaciones que se representan, y esta abstracción permite una rápida y eficiente procesamiento de las conexiones dentro de los datos.

Por otra parte, se puede mencionar a las bases de datos de grafos, que tienen una gran fortaleza en la coincidencia de patrones basada en cómo se relacionan las entidades. Los casos de uso típicos para este tipo de consulta implican cosas como

motores de recomendación, detección de fraude o detección de intrusiones (Bechberger, 2020).

Dado que cada etapa o actuación del proceso de responsabilidad fiscal, contiene atributos como fechas, números, cuantías, nombres, entre otros, los consideramos dentro del modelado de los grafos a estas actuaciones como los vértices. En consecuencia, como las actuaciones se relacionan entre sí y comparten información procedemos a denominar estas relaciones como las aristas del grafo. En el análisis de la información, encontramos que estas aristas llevan un sentido, es decir de la actuación de auto de apertura se continua con una etapa de auto de pruebas y de dicha etapa se puede ir a una actuación de decisión de la etapa de apertura, por lo que, en la esquematización de los grafos, se observa una relación dirigida, que es tenida en cuenta al momento de la creación del diagrama.

Un grafo dirigido, define el sentido en que están conectados dos vértices por una arista, se representa mediante una flecha la dirección en que se conectan dos nodos, (Grimaldi, 1998) esta misma representación se tiene en cuenta en todos los grafos esquematizados.

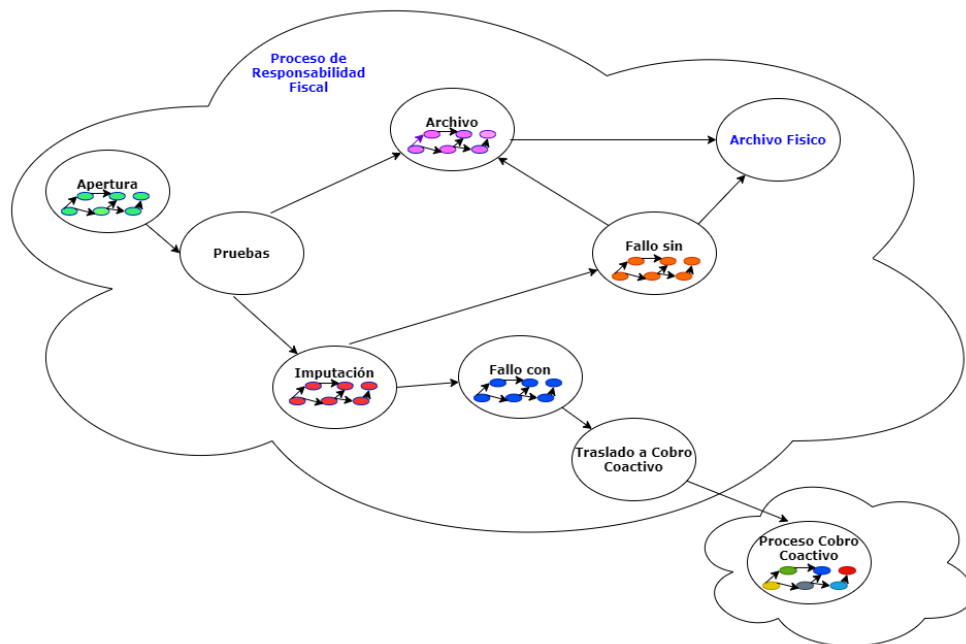
Finalmente hablaremos del valor de las aristas que se presentan en los grafos ponderados. La ponderación o peso es la asociación de un valor a una arista. [38] En las representaciones propuestas, se establece un peso y se muestra en el grosor de la arista que une dos actuaciones procesales, relacionados con la cantidad de actuaciones encontradas que visitan los vértices de los nodos diagramados. Esto nos permite

encontrar tendencias de tránsito de los expedientes con causal comprobada de prescripción. Y en estos nodos, es donde enfocaremos inicialmente el estudio para encontrar patrones de comportamiento.

6.7 Manejo del proceso de responsabilidad fiscal

Figura 2

Actuaciones del Proceso de Responsabilidad Fiscal



Fuente: Elaboración propia (2021)

Según lo determina la ley 610 de 2000 (Colombia, 2000) y lo ratifica, establece y fortalece el Decreto 403 de 2020 (Colombia, 1991), el proceso responsabilidad fiscal tienen cuatro etapas grandes

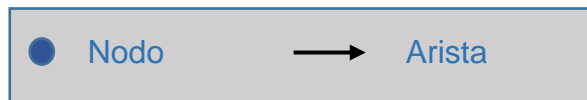
1. La apertura

2. La imputación
3. El archivo
4. El fallo (Fallo Con y Fallo Sin)

Procedemos entonces a esquematizar el proceso mediante un grafo en donde los nodos serán las etapas del proceso y las aristas las conexiones que existen entre las etapas. Para pasar de una etapa “A” a una etapa “B” se hace necesario que una arista conecte los dos nodos, en el sentido que se desea transitar. Por ello la representación de nodo y arista se hará según:

Figura 3

Representación de nodo y arista



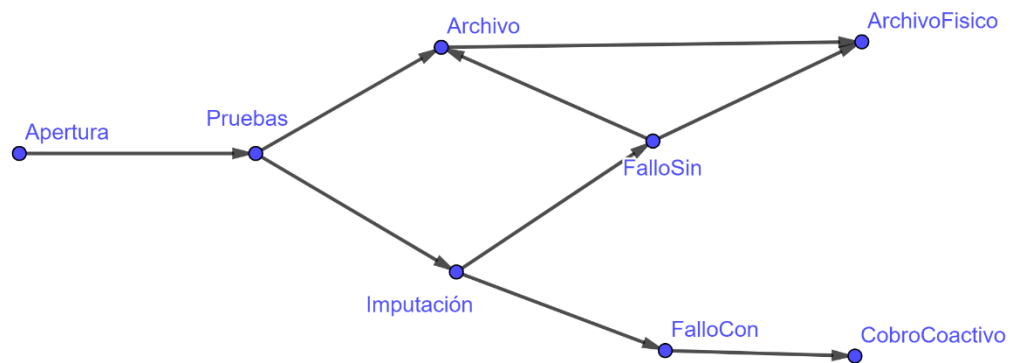
Fuente: Elaboración propia (2020)

En el siguiente gráfico se muestra el grafo general del proceso de responsabilidad fiscal. Este proceso comienza con una fase de apertura, en donde se realizan las etapas iniciales del proceso. Posteriormente con un Auto de decreto de pruebas, en donde se establecen el posterior actuar del proceso, se toma una decisión que puede ser Imputación o Archivo. Si el camino transitado es la imputación, la decisión final de esta etapa puede ser el fallo con responsabilidad fiscal, que lleva a un

proceso de cobro coactivo o el fallo sin responsabilidad fiscal que conduce a un archivo físico.

Figura 4

Grafo de actuaciones ley 610 de 2000

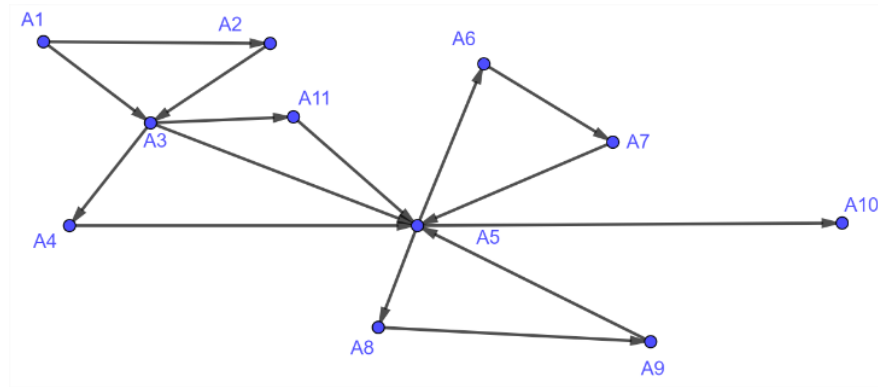


Fuente: Elaboración propia (Diseñado con GeoGebra) (2020)

Todo proceso comienza por el auto de apertura y sus correspondientes comunicaciones y/o Notificaciones reguladas por el procedimiento contencioso administrativo en sus artículos 44 y 45 (Colombia., 1992). En las pruebas decretadas en dicha etapa, puede ser que el implicado logre demostrar que no amerita una investigación profunda, por lo que el abogado sustanciador decrete el archivo. Existen caminos alternos, que no hacen parte del tránsito normal del proceso, pero que de una u otra manera pueden en casos específicos ser transitados, entre estas actuaciones tenemos la suspensión y prorrogas de los procesos. Todas estas alternancias tienen como finalidad una decisión reflejada en el siguiente grafo en la Actividad A10, que corresponde a la decisión de la etapa de apertura.

Figura 5

Grafo etapa de apertura proceso de responsabilidad fiscal



Nodo	Descripción
A1	Auto de Apertura
A2	Comunicación
A3	Notificación Auto de Apertura
A4	Nombramiento apoderado de oficio
A5	Auto que decreta Pruebas
A6	Auto que decreta la prórroga
A7	Auto que suspende la prórroga
A8	Auto de suspensión del proceso
A9	Auto reanudación del proceso
A10	Decisión de la apertura (Archivar o Imputar)
A11	Fecha comunicación vinculación al representante legal de la compañía de seguros

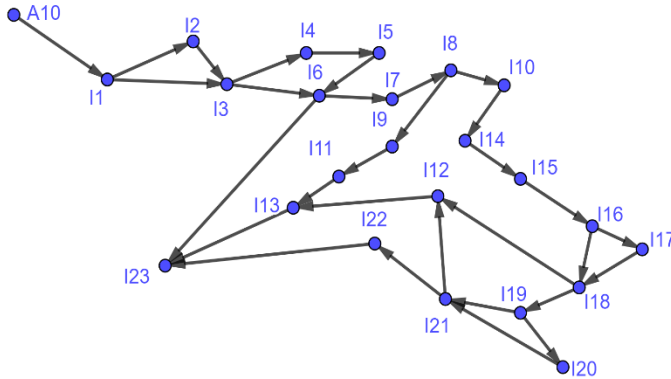
Fuente:

Elaboración propia (Diseñado con GeoGebra) (2020)

Luego de la etapa de decisión en el auto de apertura, pero siguiendo un camino que conduzca a la imputación, se da inicio a esta etapa con el Auto de imputación de cargos con sus correspondientes etapas iniciales, que al igual que las comunicaciones del auto de apertura están reglamentadas por el procedimiento administrativo, en lo que concierne a la comunicación y notificación (Colombia., 1992).

Figura 6

Grafo etapa imputación proceso de responsabilidad fiscal



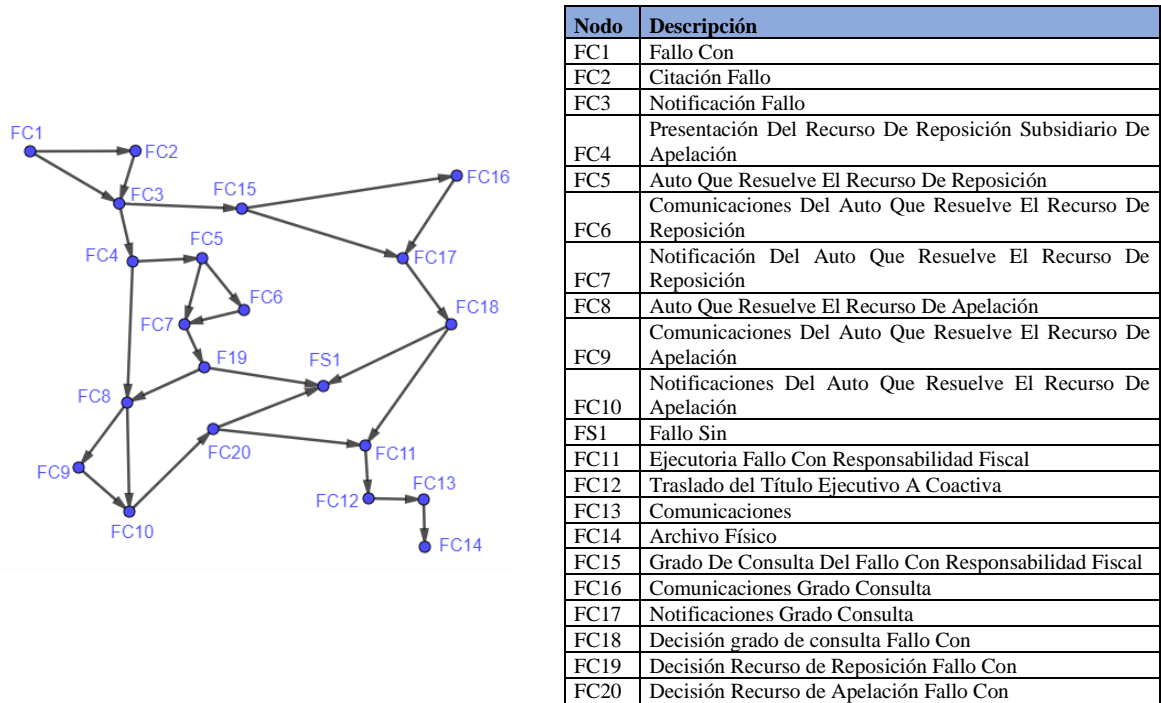
Nodo	Descripción
A10	Decisión De La Apertura (Archivar Imputar)
I1	Auto De Imputación
I2	Comunicación
I3	Notificación Auto De Imputación
I4	Nombramiento Apoderado de Oficio
I5	Notificación Personal Del Apoderado
I6	Traslado Para Descargos
I7	Presentación Descargos
I8	Auto Que Decreta O Niega Pruebas
I9	Decreto De Pruebas
I10	Negación De Pruebas
I11	Notificación Por Estado De Decreto Las Pruebas
I12	Revoca El Auto Que Niega Pruebas
I13	Practica De Las Pruebas
I14	Notificación Personal De La Negación De Pruebas
I15	Presentación De Recurso De Reposición Subsidiario De Apelación
I16	Auto Que Resuelve Recurso De Reposición
I17	Comunicaciones Del Auto Que Resuelve El Recurso De Reposición
I18	Notificación Del Recurso De Reposición
I19	Auto Que Resuelve Recurso De Apelación
I20	Comunicaciones Del Auto Que Resuelve El Recurso De Apelación
I21	Notificación Del Recurso De Apelación
I22	Confirma El Auto Que Deniega Pruebas
I23	FALLO

Fuente: Elaboración propia (Diseñado con GeoGebra) (2020)

Completando esta etapa informativa, se corre traslado para que el implicado pueda presentar sus descargos y solicite las pruebas, que considere que no se han tenido en cuenta en la etapa preliminar. Dicha etapa de pruebas tiene varios caminos dependiendo de la decisión del sustanciador, donde este último puede llegar a decretar o negar la práctica de dichas pruebas. Con el fin de garantizar el estado de derecho, se dan términos de presentación de recursos contra el actuario del sustanciador. Si estando en la etapa de imputación, el abogado que lleva el proceso, no tiene el material probatorio suficiente sobre los presuntos imputados, esto conlleva a un fallo sin responsabilidad Fiscal (Salgado., 2021).

Figura 7

Grafo etapa Fallo con Proceso de Responsabilidad Fiscal



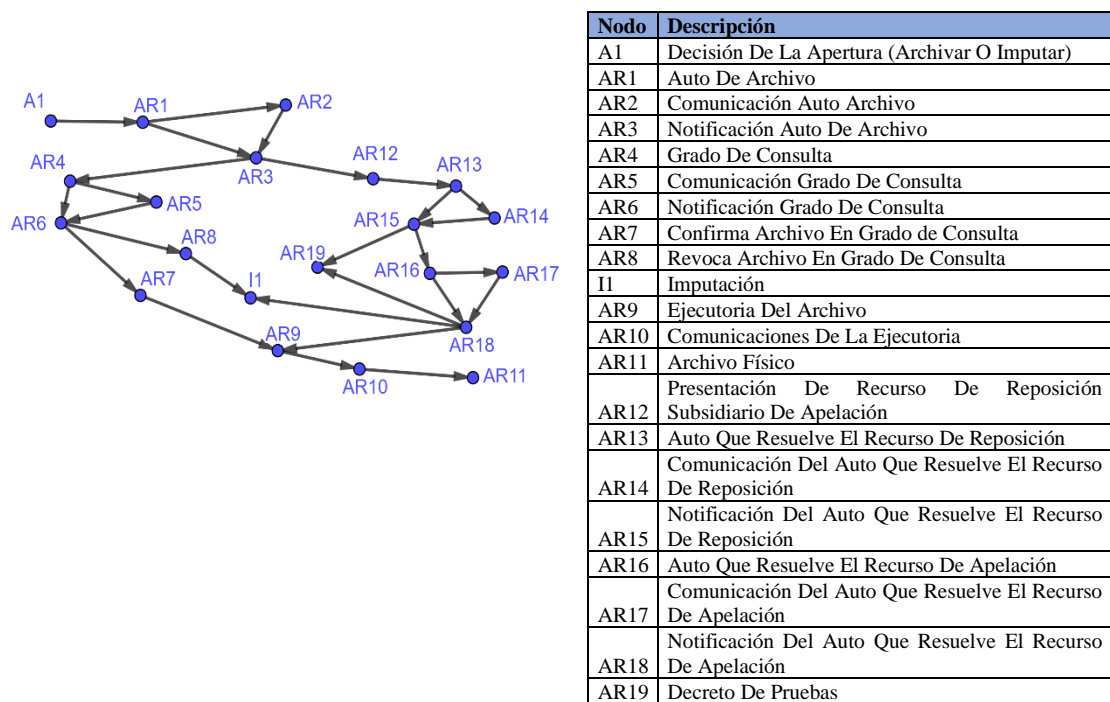
Fuente: Elaboración propia (Diseñado con GeoGebra) (2020)

Estando en la etapa de imputación y después de la práctica de pruebas y el correspondiente tránsito procedimental, es posible que el sustanciador tenga los elementos probatorios para determinar que los presuntos imputados son responsables, esto conlleva a un fallo con responsabilidad Fiscal. Esta etapa inicia con un auto que decreta el Fallo con responsabilidad fiscal, en dicho fallo, se hace un recuento del actual procesal del expediente, así como también se menciona la cuantía final y las normas infringidas por los implicados, comprobados en los hechos investigados. Al igual que en las anteriores etapas la parte informativa esta reglada por el contencioso administrativo (Colombia., 1992)

En esta etapa se incluye un actuar adicional y es el relacionado con el componente jurídico, que permite mediante la elevación del grado de consulta procesal, se subsanen las posibles inconsistencias jurídicas, que en el actuar del proceso puedan haber quedado inconclusas, incompletas o con tendencia a una libre interpretación.

Figura 8

Grafo Etapa Archivo Proceso de Responsabilidad Fiscal



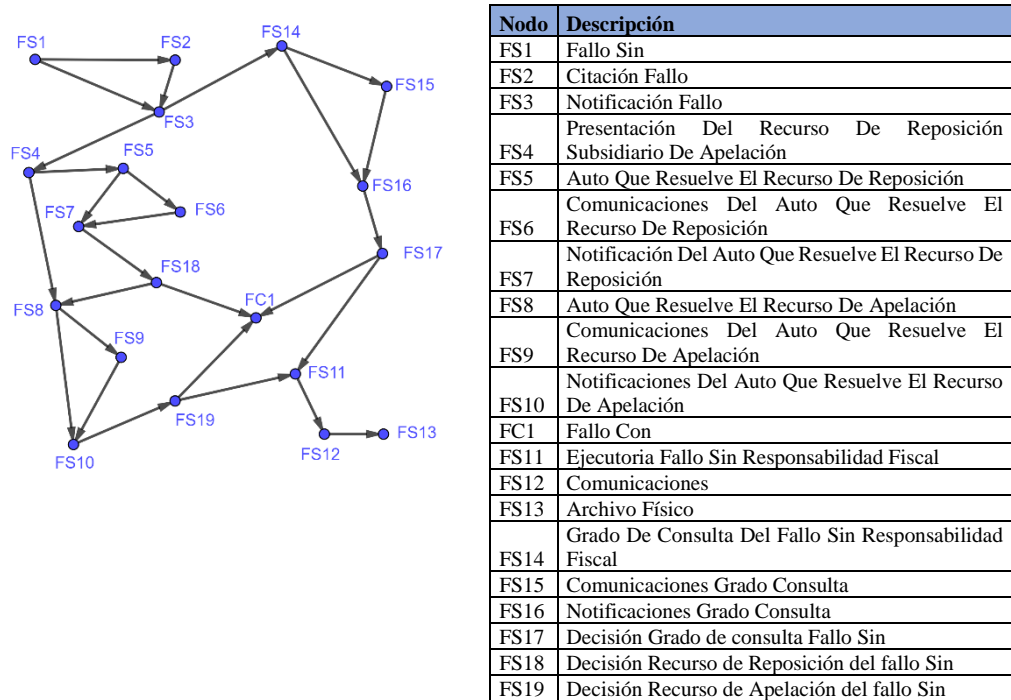
Fuente: Elaboración propia (Diseñado con GeoGebra) (2020)

Sin embargo, desde la etapa de apertura se puede tener el adecuado material para imputar cargos, pero en el transcurso de la imputación es posible que los presuntos imputados demuestren inocencia. Esto conlleva a que se decrete el Archivo. Esta etapa al igual que con las etapas anteriores, esta soportada en etapas informativas de

comunicación y notificación del auto de archivo (Colombia., 1992). Y al igual que con la etapa del fallo con responsabilidad fiscal, se tiene una etapa de revisión jurídica, por parte de la segunda instancia de la entidad, en la etapa denominada grado de consulta. De igual manera con el fin de garantizar el derecho del investigado este puede presentar los recursos correspondientes. Finalmente, y al igual que con los fallos con y sin responsabilidad la culminación de esta etapa finaliza con el auto que decreta la ejecutoria.

Figura 9

Grafo Etapa Fallo Sin Proceso de Responsabilidad Fiscal



Fuente: Elaboración propia (Diseñado con GeoGebra) (2020)

Entre todo el tránsito del proceso, luego de las pruebas decretadas en la etapa de imputación, el sustanciador puede determinar que no hay suficiente material probatorio que permita determinar la responsabilidad fiscal de los presuntos implicados o si, luego de tener un fallo con responsabilidad fiscal, los imputados demuestran que las pruebas presentadas en etapas anteriores no son concluyentes, se procede a decretar un fallo sin responsabilidad fiscal. Al igual que las anteriores actuaciones, la información contenida en el auto está regulada, para ser comunicada por el contencioso administrativo (Colombia., 1992).

Y aunque casi nunca utilizado (Salgado., 2021) también con el fin de preservar el debido proceso se tienen los recursos correspondientes. Finalmente, y al igual que en el archivo y fallo con, encontraremos el grado de consulta, suficientemente ya explicado en las etapas anteriores.

Todo Proceso con terminación anormal desemboca en los pasos del archivo.

La prescripción es una de las diversas formas de terminación anormal que se da, cuando contando 5 años a partir del auto de apertura, no se ha tomado una decisión de fondo.

Finalmente, los procesos que logran llegar a una terminación con fallo con responsabilidad fiscal son trasladados a cobro coactivo, para que allí comience una investigación que permita recuperar las cuantías adeudadas (Ordóñez, 2019).

Se observa que la dinámica del proceso es manejada como un grafo de actividades. La anterior representación es tomada de las etapas macro por las que pasa

un proceso de responsabilidad fiscal y es el punto de partida para analizar la información histórica recolectada. A medida que se avance en las etapas del proyecto, se desagregara cada una de esas etapas en aras de un mayor entendimiento procedimental y mejor acoplamiento a la técnica de inteligencia artificial que en este documento se sugiere.

6.8 Aprendizaje de máquina

La inteligencia artificial y el Machine Learning también se aplican para encontrar y combatir el lavado de dinero. Las autoridades aduaneras utilizan la IA para predecir la evasión fiscal o para monitorear e identificar licitaciones sospechosas en la contratación pública. Caso práctico es cómo Transparencia Internacional de Reino Unido está adoptando soluciones de inteligencia artificial para automatizar las búsquedas en registros públicos y fortalecer su capacidad analítica (Lima, 2020). Es importante resaltar que lo que se busca con la implementación del aprendizaje supervisado es un aumento de productividad laboral con el incremento de automatización, que permita la detección temprana de anomalías, acompañado de la reducción de los defectos en la manufactura, al ejecutar análisis de riesgos y entender los patrones de falla, por otra parte, la mejora de los productos y servicios en el análisis de estos datos garantiza mejorar la experiencia del consumidor (Gómez, 2020). Gómez et al. (2020), en este mismo artículo se resalta un caso práctico en donde mediante una aplicación para dispositivos móviles, implementada en Brasil y basada en el manejo de riesgos de incumplimiento para créditos agrícolas, se busca mediante modelos de predicción, que se mejoren las calificaciones de riesgo del agricultor y aumenten las

oportunidades de inversión. Otro de los ejemplos que se puede encontrar en la literatura es el aplicativo PREDPUL, está basado en la lógica de que el mejor predictor para un delito futuro es un delito pasado (Benbouzid, 2019).

6.9 Aprendizaje Automático

El Aprendizaje Automático es una rama de la Inteligencia artificial en donde el objetivo es lograr el aprendizaje de las máquinas. La idea es crear aplicaciones con la capacidad de generalizar el comportamiento a partir de la información estructurada. Los estudios en esta rama comprenden, los motores de búsqueda, medicina, detección de fraudes, entre otros. El resultado de este aprendizaje, normalmente son modelos matemáticos que explican el comportamiento de los datos. Estos modelos pueden ser geométricos, que pueden tener más de una dimensión, Probabilístico, los cuales determinan la distribución de probabilidades en función de las características de los datos, Lógicos, que, a través de árboles de decisión, expresan la probabilidad de un evento. Según sus funcionalidades y aplicaciones, el aprendizaje automático agrupa un conjunto de algoritmos que pueden ser clasificados en: Aprendizaje Supervisado, el cual genera una función que tiene relación de las entradas con las salidas, su uso más notorio es la clasificación de datos, en donde el algoritmo intenta marcar con una etiqueta un conjunto de vectores. Mediante los parámetros de entrada y controlados por un agente externo, quien se encarga de validar la respuesta generada, este realiza un entrenamiento al algoritmo, que permita implantar el aprendizaje deseado. La segunda clasificación y que se denomina, Aprendizaje No Supervisado, se diferencia del primero, en que las clases en las que se podría clasificar la información, se desconocen,

por lo que la misión del sistema será, encontrar los patrones que logren clasificar y etiquetar la información. Su uso está enfocado en la agrupación y la búsqueda de características relacionales de los datos de entrada. Finalmente, encontramos el aprendizaje semi-supervisado, que combina los aprendizajes ya mencionados y el aprendizaje por esfuerzo, el cual utiliza el entorno, es decir, utiliza los datos de entrada como una retroalimentación de las acciones obtenidas del mundo exterior. El aprendizaje es tomado por el método de ensayo y error (Daza, 2016).

6.10 Algoritmos de Minería de Datos

En el campo del aprendizaje supervisado, existen dos tipos de algoritmos, Clasificación y regresión. En el primero esperamos que el algoritmo logre identificar, el grupo correcto al que pertenece el registro que se encuentra estudiando. La función del algoritmo será encontrar patrones y clasificar los registros en grupos. La variable a predecir es un estado que puede ser categóricos o discretos. Por su parte el algoritmo de regresión no lo ubica en un grupo, sino que devuelve un valor específico (Sandoval, 2018).

Entre los algoritmos existentes encontramos:

A. Gradient Boosted Tree: Es utilizado el ensamble de árboles de decisión secuencial. El objetivo es minimizar una función de pérdida, optimizando los resultados obtenidos en cada árbol generado, con el fin de tener una mejora en la generación del entrenamiento de los árboles posteriores (Sandoval, 2018).

B. Random Forest: Se utilizan varios algoritmos para árboles de decisión en forma paralela, cada árbol depende de un vector que es generado de manera aleatoria. Los árboles sin correlación se promedian con el fin de obtener un resultado final. Puede resolver tanto problemas de clasificación, así como también de regresión (Daza, 2016) (Sandoval, 2018).

C. Regresión Lineal: Utiliza un modelo matemático que se encarga de encontrar una dependencia entre una variable predictiva, basado en otras independientes. El objetivo es intentar reducir el error cuadrático con una recta que descarte los residuos del modelo (Daza, 2016).

D. Árboles de Clasificación: Hace referencia a un conjunto de decisiones organizadas en una estructura jerárquica. La decisión final se determina siguiendo las condiciones que se toman desde la raíz del árbol hasta una de las hojas (Daza, 2016).

6.11 Grafos de conocimiento

Son una representación del conocimiento. Dicha técnica permite modelar entornos ricos en información histórica, en donde el grafo tiene atributos asociados a los nodos. Esto hace que el grafo simple, se convierta en una estructura que no solamente es una representación de datos, sino que también permite mediante técnicas de Machine Learning e inteligencia artificial, la generación de conocimiento. Este conocimiento está basado en estructuras matemáticas con vértices como unidades de conocimiento, conectadas por aristas que representan la relación de requisitos previos.

Este concepto es tomado del aprendizaje entre estudiantes y profesores, pero puede ser aplicado al aprendizaje y la enseñanza automática. Dicho concepto está definido como un grafo de conocimiento según lo menciona (Bellomarini, 2020) y tiene como objetivo enriquecer la representación de la estructura gráfica con el fin de manejar la complejidad del mundo real para potenciar la inferencia, el aprendizaje y razonamiento. Un ejemplo de aplicabilidad de esta técnica la tenemos con el Banco Central de Italia, en donde basándose en la detección de conexiones personales sobre gráficos de propiedad, se predicen enlaces ocultos en estructuras de red, que permiten detectar, derechos de una empresa sobre otra, vínculos de relación de acciones y vínculos personales y/o familiares (Atzeni, 2020).

6.12 Bases de datos orientadas a grafos

Un grafo es considerado como una representación binaria de entidades como nodos y cuyas relaciones entre ellos a través de líneas denominadas aristas. Permiten que se estructuren los datos de forma flexible y con variación de relaciones. Por otra parte, podemos establecer relaciones de precedencia más generales que en las listas y los árboles (Pinilla, 2017).

Cuando nos referimos a una base de datos orientada a grafos, nos referimos a una base de datos NoSQL. En estas bases de datos orientadas a grafos se utilizan los nodos y las aristas de un grafo para representar datos sus relaciones de tal manera que la teoría de grafos puede ser utilizada para recorrer la base de datos (Calabuig, 2018) (Solé, 2018).

6.13 Implementación en tramas de corrupción

Una de las bases de datos orientadas a grafos es Neo4j, Este se utilizó para la detección de grandes tramas de corrupción. Su uso ha sido llevado en temas como la lista de Falciani, los Paradise Papers y los Papeles de Panamá, entre otros (Calabuig, 2018). En los papeles de Panamá a partir de ficheros de texto, en donde contenían los datos y las relaciones con las empresas, oficinas y clientes que formaban parte del entramado, se creó un grafo que permitía la comprensión de las interacciones entre los diferentes actores de esta red. Ya en el servidor se hace una búsqueda BFS (Primero en Amplitud) partiendo desde el nodo seleccionado hasta llegar al nivel de profundidad requerido. Los vértices y aristas se van guardando para devolverlos como resultado (TodoBI, 2021).

Otro de los ejemplos encontrados en la literatura y cuya implementación está basada en la teoría de grafos, es la realizada por Abarca Jorge (Abarca Aguirre, 2010), en donde explica la investigación del mexicano Juan Castaingts Teyllery, sobre el estudio de la corrupción y quien señala que, “desde un punto de vista analítico, no interesa calificar si alguna persona o institución son o no corruptos, lo que interesa a la metodología es si tiene

o no relaciones corruptas”. Para ello mediante el uso de la teoría de grafos, suministra el instrumento para estudiar un sistema corrupto integrado por cuatro individuos, entre los que se establecen un conjunto de relaciones corruptas y no corruptas. Los vértices son representados por puntos y las aristas por líneas. Una relación corrupta es

representada con el signo algebraico positivo (+) y una relación no corrupta por el símbolo negativo (-). De esta manera la presencia de símbolos negativos, es entendida como ausencia de corrupción. Por su parte, si predominan las relaciones negativas se dice que se trata de un sistema corrupto inestable. El criterio de operación es similar a la regla con la que funcionan los signos del álgebra normal, en la siguiente figura se esquematiza el correspondiente uso:

Figura 10

Operación de relaciones

$(+) \times (+) = +$	Multiplicación de símbolos iguales da como resultado positivo.
$(-) \times (+) = -$	Multiplicación de símbolos diferentes da como resultado negativo.

Fuente: Elaboración propia (2021)

Finalmente, el autor muestra con diversos grafos las respectivas comprobaciones que le permiten la definición de sistemas corruptos estables y no estables.

En Colombia se tiene como ejemplo un estudio realizado por la fundación VORTEX para la Procuraduría General de la Nación, en donde lo que se busca es mediante el uso de la teoría de grafos, poder identificar redes de macro corrupción. El estudio se basa en poder determinar cómo nodos del sistema, actores legales e ilegales que establecen múltiples y diversas formas de interacción, crean extensas redes ilícitas que se modelan y analizan para entender la dimensión, la complejidad, y los efectos a niveles institucionales del fenómeno en los entes gubernamentales. Mediante la

información cualitativa almacenada en el Sistema Información Misional (SIM) de la Procuraduría General de la Nación, complementada con información física de los expedientes actuales, posteriormente sistematizada para su análisis, así como también investigaciones finalizadas de procesos disciplinarios culminados de investigaciones de irregularidades en programas de alimentación escolar en la Guajira, Caquetá y Putumayo. Todas estas investigaciones basadas en el protocolo de Análisis de Redes Criminales de la Procuraduría General de la Nación (ARCPGN). Dicho protocolo está basado en el análisis de redes criminales que a su vez se basa en el análisis de redes sociales (Vortex, 2020).

A nivel nacional también podemos hablar de la implementación de la central de información contractual OCEANO en la Contraloría General de la Nación. Proyecto de explotación de datos emprendido entre el ocaso de 2018 e inicio de 2020, basado en bases de datos internas y actualizadas de la propia entidad y acceso a sistemas de información externos. El objetivo es realizar vínculos y detección de patrones de comportamiento en la contratación pública. Esto apalancado en el uso de la analítica de datos, las mallas de información y la inteligencia artificial. Utilizando algoritmos de malla basados en la teoría de grafos y en conceptos tales como cercanía (distancia entre nodos), grado (número de enlaces que posee un nodo) e intermediación (número de veces que un nodo actúa como puente) y el uso de algoritmos de predicción. Se intenta detectar a los fallecidos que firman contratos, contratistas relacionados en colusión (Acuerdo entre dos o más partes para limitar la competencia), inhabilitados

(sancionados que según la ley no pueden contratar con el estado) y con investigaciones de responsabilidad fiscal (Sánchez, 2020) (CGR, 2019).

6.14 Balanceo de datos

Dado que los datos suministrados por la Auditoría General de la República, contienen una cantidad alta en procesos de responsabilidad fiscal en curso y procesos con terminaciones normales, se hace necesario para el análisis de la información, encontrar una técnica de balanceo de datos, que permita tomar una muestra de entrenamiento acorde con la investigación, es decir una muestra representativa, que contenga procesos con terminación por prescripción. Teniendo en cuenta la baja representación de procesos con terminación anormal y más específicamente con terminación por prescripción, se procede a indagar una técnica de minería de datos que nos permita equilibrar la muestra a tomar para el entrenamiento del algoritmo de clasificación.

Según la revisión de literatura se encuentran varias técnicas, entre ellas las más representativas:

- a. Submuestreo
- b. Sobremuestreo
- c. Smote
- d. Variaciones de Smote

Cada una de ellas con las correspondientes limitaciones (Asniar, 2021). En submuestreo (undersampling), se toma la clase minoritaria como techo de los datos a tomar, para lograr un equilibrio entre clases, con la desventaja de escasos datos y posible pérdida de datos relevantes. En el sobremuestreo (oversampling) en cambio, se duplican los datos de la clase minoritaria, con el fin de generar un sobreajuste que permita acercar los datos al techo de la clase mayoritaria, con el riesgo de provocar sesgos hacia la clase mayoritaria (Rodríguez, 2009). Por otra parte, Smote se basa en el sobremuestreo, pero generando instancias artificiales con variaciones en los atributos, interpolando mediante el ayuda del azar o métodos determinísticos, cada muestra minoritaria con el vecino más cercano de dicha clase. La principal desventaja es el alto coste computacional que tiene (Sposito, 2020) (Asniar, 2021) (Arrieta, 2015).

Finalmente, a partir de Smote, se han desarrollado diversos algoritmos basados en este, que mejoran el balanceo y rendimiento en el aprendizaje desequilibrado (Torres-Vásquez, 2019).

7. ESTRATEGIA METODOLÓGICA

Figura 11

Estrategia Metodológica



Fuente: Elaboración propia (2021)

En cuanto a la estrategia metodológica a utilizar para el proyecto, partiendo de la premisa de que el proceso de responsabilidad en las contralorías territoriales contiene información digital desde hace varios años y que dicha información es reportada a los entes de auditoría de la nación, procederemos a establecer una metodología cuantitativa que iniciará por una recolección de material digital, mediante la solicitud ciudadana del derecho de petición y la recolección de contactos que permiten lograr reuniones y entrevistas con los directos responsables. Luego de obtener la información y de proceder a realizar una valoración de relevancia de la misma, se establecerán hipótesis, definición y uso de las variables implicadas en la información, esto con el fin de ser utilizadas en etapas posteriores. Después se realizará un análisis completo de la información recolectada y aplicando técnicas de extracción y manejo de la información, como minería de datos o aprendizaje de máquina, procederemos a

analizar, diseñar y desarrollar un algoritmo que permita comprender y detectar el comportamiento de los procesos de responsabilidad fiscal que llevan la marca de la prescripción. Esto permitirá en la siguiente etapa, validar los resultados, que permitan aplicar un desarrollo complementario, que detecte los procesos actuales y muestre de manera preventiva, según la predicción, los procesos en curso que están en riesgo de una causal de terminación anormal por el fenómeno investigado. Finalmente se presentarán los resultados y mencionarán las acciones de mejora que permitan una precisión más acertada del algoritmo.

8. CRISP-DM (Cross Industry Standard Process for Data Mining)

Figura 12

Metodología CRIS-DM



Fuente: Elaboración propia (2021)

En cuanto a la metodología de desarrollo del proyecto se pretende utilizar la metodología CRISP-DM. Esto a razón de que es una metodología de una alta confiabilidad, probada en la exploración y visualización de datos para descubrir patrones sospechosos en datos financieros. CRISP-DM permite crear un modelo de minería de datos que se adapte a sus necesidades concretas (IBM, 2020). Basándose en cada una de las etapas que se proceden a enunciar y sustentar a continuación:

A. Comprensión del negocio o problema:

En esta etapa se establecen las tareas de comprensión de objetivos y los requisitos del proyecto, los objetivos técnicos y un plan de proyecto. En aras de sacar el mejor provecho de la minería de datos, es primordial entender el problema a resolver, esto nos lleva a recolectar los datos correctos e interpretar adecuadamente los resultados. En esta fase, es importante convertir el conocimiento adquirido del negocio, en un problema que tenga relación con la minería de datos, esto con el fin de crear un plan previo, cuya meta finalmente sea poder alcanzar los objetivos del negocio (IBM, 2020).

La idea es que se utilice esta etapa de la metodología para entender con los responsables de la información, lo relacionado a las minucias procesales y las eventualidades presentadas en los procesos con terminaciones anormales, teniendo en cuenta la determinación de los objetivos del negocio, la evaluación de los mismos y representar la finalidad de la implementación del algoritmo basado en grafos para posteriores etapas. Por otra parte, en esta se procederá a verificar la legislación actual relacionada a los procesos de responsabilidad fiscal en Colombia.

B. Fase de comprensión de los datos

Establecer un contacto inicial con el problema con el objetivo de comprender adecuadamente el problema. Para esta fase estableceremos las primeras hipótesis y se gestionará un proceso más cercano a los datos que maneja el negocio, con el fin de entender los sistemas con que cuentan las contralorías territoriales y que están relacionados, de forma directa o indirecta con el manejo

de los expedientes de las entidades de control en Colombia. Para ello mediante derechos de petición se pedirán a las entidades una relación de las herramientas tecnológicas y que contengan información histórica de los procesos de responsabilidad fiscal. Otra de las opciones que se tuvieron en cuenta en esta etapa, corresponde a la posibilidad de implementación del proyecto en las etapas propuestas por el SINACOF, encontrando que, a raíz de que la implementación propuesta abarca los procesos de responsabilidad fiscal que pueden ser afectados por la prescripción y no en la mejora de tiempos de ejecución de las actividades.

En este proyecto se desestima tener en cuenta la implementación del mismo como mejora o aporte significativo al SINACOF.

C. Fase de preparación de los datos

El objetivo de esta fase será extraer y preparar los datos con el fin de realizar los análisis que permitan encontrar las relaciones entre variables y selección específica de los datos que servirán para hacer los estudios de seguimiento a los procesos con terminación anormal. También se procederá a realizar una depuración de la información en aras de desechar registros incompletos o que genere ruido a desnivel de las métricas inicialmente recolectadas en fases anteriores. De igual manera se estructurará la base de datos que permita albergar y realizar las pruebas de testeo de la información. Mediante consulta a la Auditoría General de la República, se obtuvo información de los procesos de responsabilidad fiscal desde el año 2010 hasta la última vigencia fiscal que fue el 2019.

D. Fase de Modelado

En esta etapa, con la información recolectada se propone analizar la base de datos mediante un algoritmo y se procederá a escoger una técnica de modelado que permita extraer la información comparativa con los procesos activos y sobre los cuales queremos mitigar la probabilidad de repetición. Se construirá un plan de pruebas y se evalúan los resultados con el fin de ajustar las variables, para de esta manera lograr una mayor precisión (IBM, 2020).

E. Fase de evaluación

Con los resultados obtenidos de la fase anterior podremos evaluar la eficiencia de los resultados y llevar a usuarios finales el desarrollo de un modelo más amigable y que permita generar la valoración de riesgo de los procesos en curso, plasmándolos en una estrategia de presentación desarrollado en Big Data, esto a razón que, sin una representación adecuada de los resultados, el conocimiento derivado es inútil (Janev, 2020). De igual manera poder sugerir mejoras y usos específicos de la herramienta.

F. Fase de implementación

Se presentarán los resultados finales para que el usuario final evalúe la calidad de la información obtenida y pueda tomar las decisiones o correctivos que se consideren. Se propondrán estrategias de monitoreo y apalancamiento de herramientas que aprovechen el nuevo conocimiento adquirido. Finalmente se documentarán las experiencias y propuestas para el mejoramiento de la herramienta (IBM, 2020).

9. DESARROLLO E IMPLEMENTACIÓN

9.1. Análisis de la información

La fuente de la información es el aplicativo SIA-SIREL de la Auditoría General de la República. Los años tomados para el presente estudio fueron 2010 a 2019, dividido por año en 10 archivos en formato XLSX. Las columnas suministradas en cada archivo fueron 49. En la siguiente tabla se discrimina la cantidad de registros suministrados en cada uno de los archivos, dichos archivos para mayor facilidad del manejo de la información fueron condensados en uno solo para un total de 144.562 registros.

Tabla 1

Registros por año Información SIA-SIREL

AÑO	REGISTROS
2010	13.707
2011	14.884
2012	15.002
2013	15.377
2014	15.701
2015	14.644
2016	13.226
2017	13.589
2018	14.137
2019	14.296

Fuente: Análisis propio de la información suministrada por la Auditoría General de la República

En el análisis se encuentra información de 66 contralorías entre departamentales, territoriales y distritales, así como también la Contraloría General de la República, para un total de 66 fuentes de expedientes procesales. En el anexo 1 se observa el discriminado de cada una de las fuentes.

Al condensar en un solo archivo la totalidad de registros suministrados, se adiciono una columna denominada año, con el fin de preservar el año en que se reportó el proceso. De igual manera se descartaron algunas columnas que no contenían información relevante para el estudio. Al igual que se detectó que los registros pertenecientes a la Contraloría general de la república producían un outlier en el análisis de la información, por lo que se procedió a descartar dichos registros en aras de lograr un modelo aplicable a las contrarias municipales, departamentales y distritales.

Es de aclarar que la contraloría general de la república tiene una jurisdicción nacional que según el decreto 403 de 2020 el control fiscal de esta prevalecerá cuando las investigaciones conduzcan a la perdida de recursos que la nación transfiera a las entidades territoriales (MinJD. Decreto, 2020).

Esto conlleva a que no se pueda establecer con exactitud a que contraloría territorial pertenece el proceso que se está analizando.

En cuanto a los registros que cumplen con la condición de procesos prescritos tenemos lo siguiente:

Se encontraron 498 registros por prescripción, discriminados por años, de la siguiente manera:

Tabla 2

Número de expedientes con causal de prescripción discriminados por año.

Año	Número de Expedientes
2010	18
2011	17
2012	288
2013	12
2014	9
2015	10
2016	40
2017	38
2018	38
2019	28
Total	498

Fuente: Análisis propio de la información suministrada por la Auditoría General de la República

Por contraloría se encontraron los siguientes registros

Tabla 3

Procesos de Responsabilidad Fiscal con causal de prescripción por contraloría

Contraloría	Número de Expedientes Prescritos
Antioquia	8
Barrancabermeja	1

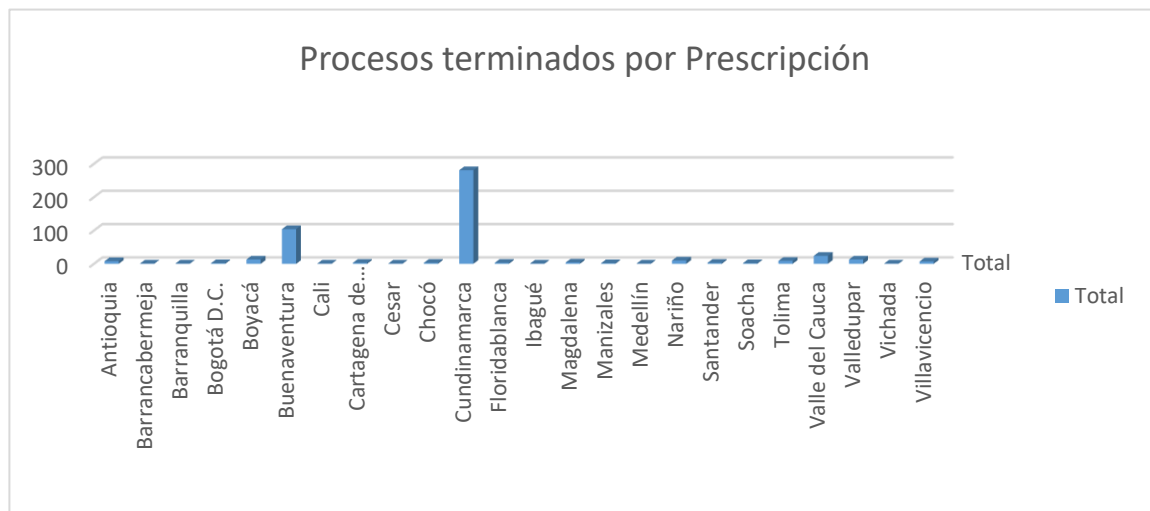
Barranquilla	1
Bogotá D.C.	2
Boyacá	13
Buenaventura	104
Cali	1
Cartagena de Indias	3
Cesar	1
Chocó	3
Cundinamarca	281
Floridablanca	3
Ibagué	1
Magdalena	4
Manizales	2
Medellín	1
Nariño	10
Santander	3
Soacha	2
Tolima	9
Valle del Cauca	24
Valledupar	13
Vichada	1
Villavicencio	7
Total	498

Fuente: Análisis propio de la información suministrada por la Auditoria General de la Republica

Según se observa, encontramos que las contralorías que reportan más casos de prescripción entre los años 2010 y 2019 son la de Buenaventura y la de Cundinamarca.

Figura 13

Procesos terminados por Prescripción Procesos de Responsabilidad Fiscal Registrados en el Aplicativo SIA-SIREL



Fuente: Análisis propio de la información suministrada por la Auditoría General de la Republica

Por otra parte, al realizar un análisis de las cuantías de procesos prescriptos encontramos lo siguiente:

Tabla 4

Cuantías de procesos prescritos

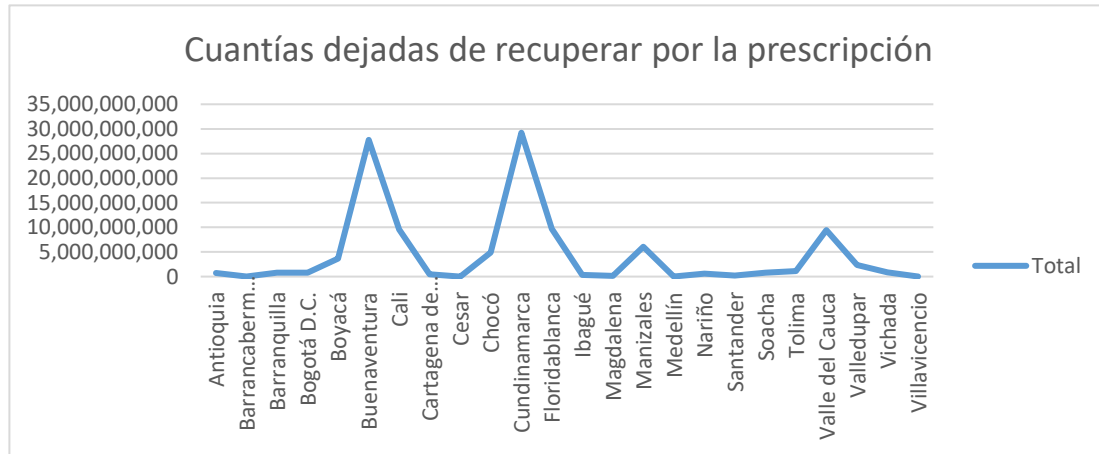
Contraloría	Cuantías dejadas de recuperar por prescripción
Antioquia	752.800.925
Barrancabermeja	45.303.387

Barranquilla	800.000.000
Bogotá D.C.	818.445.711
Boyacá	3.659.376.049
Buenaventura	27.743.133.302
Cali	9.555.639.239
Cartagena de Indias	468.269.081
Cesar	2.580.000
Chocó	4.860.569.265
Cundinamarca	29.224.361.166
Floridablanca	9.686.257.895
Ibagué	350.725.000
Magdalena	168.497.286
Manizales	6.077.867.033
Medellín	11.974.300
Nariño	617.131.027
Santander	205.926.396
Soacha	827.117.183
Tolima	1.120.121.092
Valle del Cauca	9.412.721.374
Valledupar	2.347.200.916
Vichada	870.000.000
Villavicencio	35.061.761
Total general	109.661.079.388

Fuente: Análisis propio de la información suministrada por la Auditoria General de la Republica

Figura 14

Cuántías dejadas de recuperar por la Prescripción Procesos de Responsabilidad Fiscal.

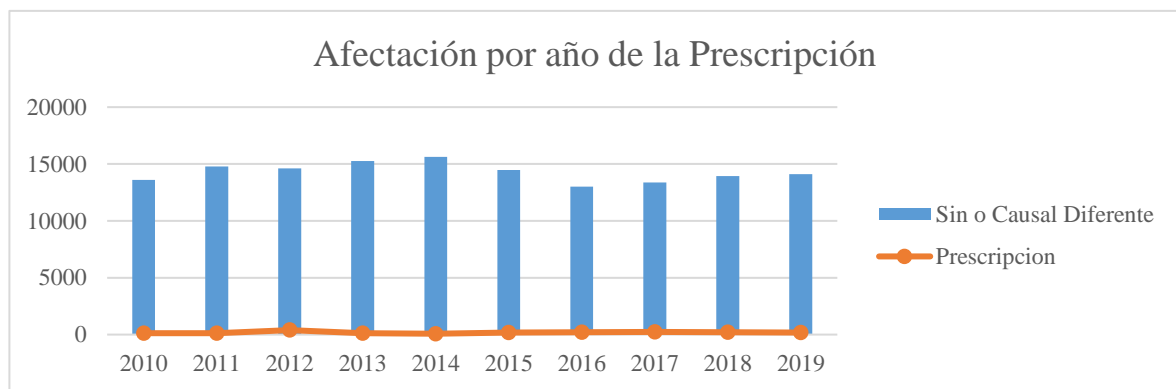


Fuente: Análisis propio de la información suministrada por la Auditoria General de la Republica

Como se observa en la gráfica anterior los mayores problemas son representados por las contralorías de Cundinamarca y Buenaventura.

Figura 15

Afectación por año de la prescripción Procesos de Responsabilidad Fiscal



Fuente: Análisis propio de la información suministrada por la Auditoria General de la Republica.

Sin embargo, al estudiar la problemática de la prescripción desde una generalidad de todos los procesos, en donde se encuentran terminaciones y otras causales ajenas a la investigación, encontramos que la afectación de los procesos con terminación anormal, asociada a la prescripción en Colombia corresponde a un 1.269% en promedio en los últimos 10 años. Dicha variación ha tenido repuntes importantes en el año 2012 y 2016. No ha tenido un descenso significativo en los últimos años. Sin embargo, las cuantías dejadas de percibir por este fenómeno han sido en los últimos 10 años por un valor de \$675.838.695.805. Esto equivale al 0.248% del presupuesto asignado para la nación en el año 2020. Valor que fue presupuestado por un valor de 271.7 billones de pesos.

Ahora bien, aunque parece un porcentaje extremadamente bajo con respecto a los procesos que se están llevando en todo el país y que pueden llegar a la recuperación del erario público en cuantías mucho mayores a las establecidas en los procesos con causales de prescripción, se observa que no todos los procesos de responsabilidad fiscal que, con auto de apertura, conllevan a un fallo favorable. Por otra parte, los procesos fallados y que corren traslado al cobro coactivo, no siempre son garantía de recuperación de los recursos, esto a raíz de diversas razones particulares de dicho proceso.

Y para finalizar y en aras de no tomar como superficial el valor de la cuantía de los procesos fiscales que termina con prescripción es importante mencionar que este 0.248% equivale a un poco más de 4 veces el valor del presupuesto asignado a la Contraloría de Bogotá para el año 2020. Valor que está en \$167.716.431.000.

9.2 Resultados en el análisis con Grafos

Para lograr un análisis general del proceso de responsabilidad fiscal y poder establecer focos específicos en donde sucede históricamente la prescripción de los procesos, se procede a tomar realizar una matriz de adyacencia que permita mediante Python y la implementación de librerías como NetworkX y Matplotlib, poder establecer gráficamente donde era más recurrente en todo el grafo del proceso, la aparición de este fenómeno.

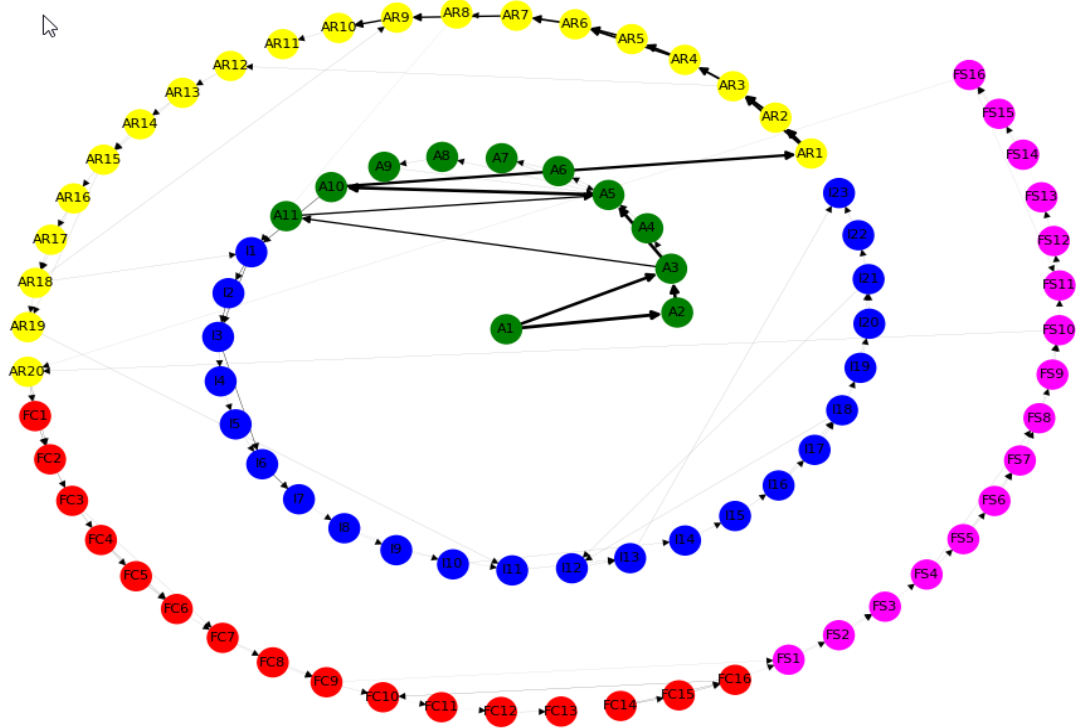
Para ello se procede a diagramar los grafos de apertura, imputación, archivo, fallo con responsabilidad fiscal y fallo sin responsabilidad fiscal, en un solo grafo, etiquetando cada vértice con el nombre asignado en la diagramación de los grafos con GeoGebra, es decir los grafos de apertura, tendrán una connotación A compuesta por $\{A1, A2, A3, A4, \dots\}$, los de imputación con I, compuesta por $\{I1, I2, I3, I4, \dots\}$, los de archivo con AR, compuesta por $\{AR1, AR2, AR3, AR4, \dots\}$, los de fallo con FC $\{FC1, FC2, FC3, FC4, \dots\}$ y finalmente los de fallo Sin $\{FS1, FS2, FS3, FS4, \dots\}$.

De igual manera se le asigna un peso a cada arista con el fin identificar los caminos más transitados. Esto se logra aumentando el grosor de la conexión entre los vértices. A mayor grosor de la arista, más procesos que transitaron por el camino. Y para identificar visualmente a que etapa pertenece cada vértice se decidió darle un color a cada gran etapa del proceso.

En el siguiente grafo se puede observar el resultado.

Figura 16

Grafo con los procesos por prescripción encontrados en SIA-SIREL.



Fuente: Análisis propio de la información suministrada por la Auditoría General de la Republica.

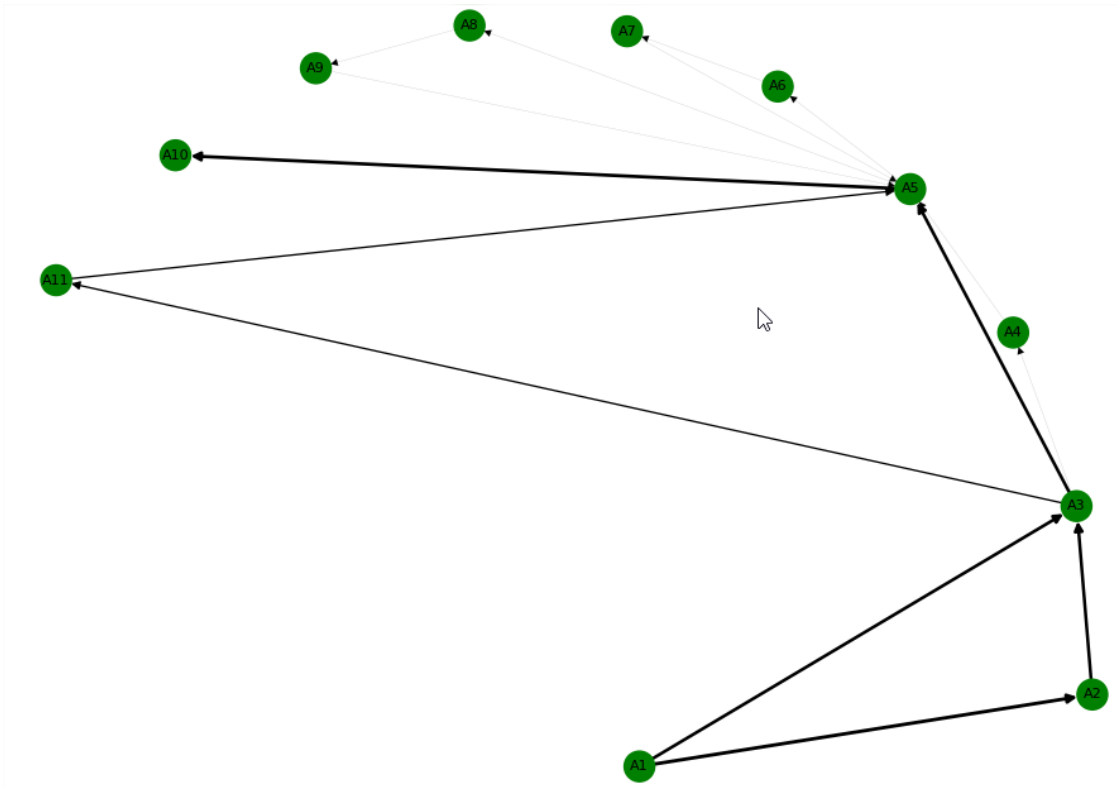
Como se puede observar en este grafo el transito más recurrente, es el encontrado en las etapas de apertura, archivo y en menor medida imputación. Por lo que se observa la necesidad de ampliar un poco más estas etapas, en aras de encontrar las actuaciones procesales más relevantes que nos permitan hallar patrones de comportamiento de la información y de esta manera lograr identificar las columnas con información importante a tener en cuenta en la creación del modelo de predicción.

Esto nos lleva a la necesidad de generar dos grafos más, que se enfoquen en las actuaciones procesales más transitadas, esto con el fin de encontrar patrones de

importancia que sirvan de guía para la implementación del modelo de Machine Learning. Uno de ellos es la apertura. El segundo es la apertura, el archivo y la imputación.

Figura 17

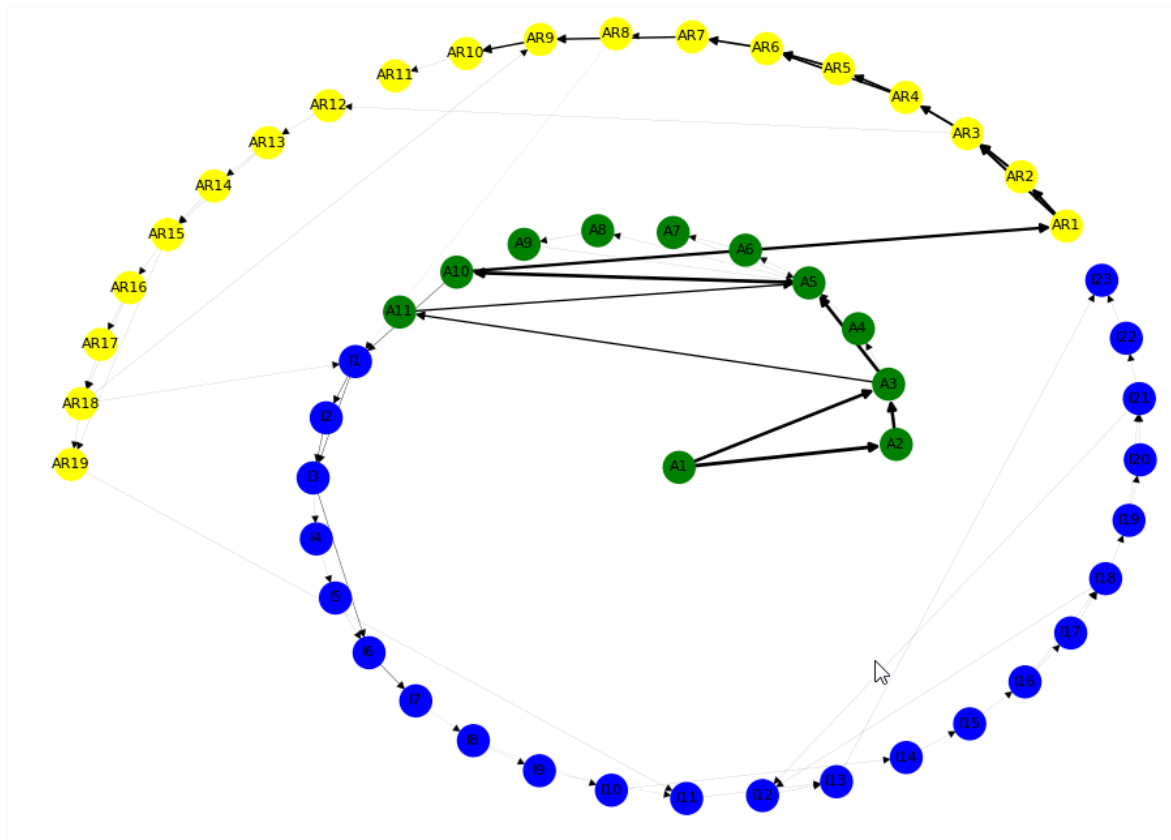
Grafo etapas de apertura procesos con prescripción encontrados en SIA-SIREL.



Fuente: Análisis propio de la información suministrada por la Auditoría General de la República.

Figura 18

Grafo etapas de apertura, archivo e imputación de procesos con prescripción encontrados en SIA-SIREL.



Fuente: Análisis propio de la información suministrada por la Auditoría General de la República.

Realizando una correspondencia de las actuaciones procesales que tenemos como propias del proceso y esquematizadas en los grafos anteriores, contra las columnas existentes en la información proporcionada en el aplicativo de la Auditoría General de la República, encontramos una correlación en la siguiente tabla en donde tomamos las actuaciones de la apertura.

Tabla 5*Concordancia de actuaciones de apertura con columnas SIA-SIREL*

Vértice según las actuaciones procesales		Columna con correspondencia en SIA-SIREL
A1	Auto de Apertura	SI
A2	Comunicación	NO
A3	Notificación Auto de Apertura	NO
A4	Nombramiento apoderado de oficio	SI
A5	Auto que decreta Pruebas	NO
A6	Auto que decreta la prorroga	NO
A7	Auto que suspende la prorroga	NO
A8	Auto de suspensión del proceso	NO
A9	Auto reanudación del proceso	NO
A10	Decisión de la apertura (Archivar o Imputar)	SI
A11	Fecha comunicación vinculación al representante legal de la compañía de seguros	SI

Fuente: Análisis propio de la información suministrada por la Auditoría General de la República

En la siguiente tabla tomamos las del archivo.

Tabla 6*Concordancia de actuaciones del archivo con columnas SIA-SIREL*

Vértice según las actuaciones procesales	Columna con correspondencia en SIA-SIREL	
A1	Decisión De La Apertura (Archivar O Imputar)	SI
AR1	Auto De Archivo	SI
AR2	Comunicación Auto Archivo	NO
AR3	Notificación Auto De Archivo	NO
AR4	Grado De Consulta	SI
AR5	Comunicación Grado De Consulta	NO
AR6	Notificación Grado De Consulta	NO
AR7	Confirma Archivo En Grado de Consulta	SI
AR8	Revoca Archivo En Grado De Consulta	SI
I1	Imputación	SI
AR9	Ejecutoria Del Archivo	SI
AR10	Comunicaciones De La Ejecutoria	SI
AR11	Archivo Físico	NO
AR12	Presentación De Recurso De Reposición Subsidiario De Apelación	SI
AR13	Auto Que Resuelve El Recurso De Reposición	SI
AR14	Comunicación Del Auto Que Resuelve El Recurso De Reposición	NO

AR15	Notificación Del Auto Que Resuelve El Recurso De Reposición	NO
AR16	Auto Que Resuelve El Recurso De Apelación	SI
AR17	Comunicación Del Auto Que Resuelve El Recurso De Apelación	NO
AR18	Notificación Del Auto Que Resuelve El Recurso De Apelación	NO

Fuente: Análisis propio de la información suministrada por la Auditoría General de la República

Finalmente, en la siguiente tabla hacemos la correspondencia para las actuaciones de la imputación.

Tabla 7

Concordancia de actuaciones de la imputación con columnas SIA-SIREL

	Vértice según las actuaciones procesales	Columna con correspondencia en SIA-SIREL
I1	Auto De Imputación	SI
I2	Comunicación	NO
I3	Notificación Auto De Imputación	NO
I4	Nombramiento Apoderado de Oficio	NO
I5	Notificación Personal Del Apoderado	NO
I6	Traslado Para Descargos	SI
I7	Presentación Descargos	NO

I8	Auto Que Decreta O Niega Pruebas	SI
I9	Decreto De Pruebas	NO
I10	Negación De Pruebas	NO
I11	Notificación Por Estado De Decreto Las Pruebas	NO
I12	Revoca El Auto Que Niega Pruebas	NO
I13	Practica De Las Pruebas	NO
I14	Notificación Personal De La Negación De Pruebas	NO
I15	Presentación De Recurso De Reposición Subsidiario De Apelación	SI
I16	Auto Que Resuelve Recurso De Reposición	SI
I17	Comunicaciones Del Auto Que Resuelve El Recurso De Reposición	NO
I18	Notificación Del Recurso De Reposición	NO
I19	Auto Que Resuelve Recurso De Apelación	SI
I20	Comunicaciones Del Auto Que Resuelve El Recurso De Apelación	NO
I21	Notificación Del Recurso De Apelación	NO
I22	Confirma El Auto Que Deniega Pruebas	NO
I23	FALLO	SI

Fuente: Análisis propio de la información suministrada por la Auditoria General de la Republica

Posterior a estas validaciones se evalúa si con la información existente en SIA-SIREL y la calidad de la misma, es suficiente para poder llegar a realizar los análisis planteados en los objetivos del proyecto. En esta etapa no sobra mencionar que estos resultados previos fueron compartidos en modalidad de Poster en el tercer congreso internacional Tic para la amazonia, realizado en la ciudad de Florencia Caquetá en noviembre de 2021. [Anexo 2]

9.3 Validación de Reglas del negocio y limpieza de datos

Como condiciones del negocio debemos tener en cuenta lo siguiente:

1. Anteriormente se mencionó que para el presente estudio se eliminaban los registros pertenecientes a la Contraloría General de la Republica lo que nos excluye 46.344 registros, dado que no se pueda establecer con exactitud a que contraloría territorial pertenece el proceso que se está analizando, así como también los valores atípicos (outliers) ya mencionados.
2. Se excluyen los registros que en el cálculo de fechas den valores negativos dado que los registros están presentados en orden cronológico. Lo que traduce como errores de digitación del expediente o nulidades procesales, que para el presente estudio no serán tenidos en cuenta.
3. De igual manera se incluye para el análisis la Base de datos de Departamentos y municipios con los códigos DANE y Regiones, disponible en Datos Abiertos (DANE, 2021).

4. Se crean en la Base de Datos original 4 columnas denominadas ID_ZONA, Zona, ID_COD_DANE, Municipio, con la columna contraloría se hace un cruce de información que permita actualizar en la base de datos original las regiones y zonas a las que pertenece cada registro.

5. Ha de tenerse en cuenta que todos los registros a analizar deben tener columnas obligatoriamente diligenciadas. Contraloría (origen de la contraloría que lleva el proceso), Entidad Afectada, Fecha ocurrencia del hecho (Fecha en la se materializa el detrimento patrimonial), Origen (forma en la que nace el proceso en la contraloría), Fecha auto de apertura PRF (Fecha en la que se da apertura al proceso de responsabilidad fiscal en el ente de control) (AGR, 2019). Los registros que inicialmente no cumplen con esta obligatoriedad, serán descartados desde el inicio para evitar inexactitud y confusión en el procesamiento de los datos.

6. El total de días que puede durar un proceso desde su auto de apertura hasta su decisión de archivo es de 5 años, después de ese tiempo el proceso sin importar la etapa en que se encuentre caerá en la causal de archivo por prescripción (Colombia, 2000).

Teniendo en cuenta que no se tienen registros documentales conocidos entre los años 2010 y 2019 que hablen de suspensión de términos, se asume la ausencia de dichas suspensiones para la creación del modelo y se tomarán 1825 días calendario (365 días de un año multiplicadas por 5) para el desarrollo total de

cada expediente. Este valor se tendrá en cuenta para las actuaciones procesales totales de cada registro analizado y de esta manera determinar, mediante el uso de cálculos sencillos tomados mediante fórmulas de Excel, los días transcurridos desde la fecha de apertura hasta la última actuación procesal registrada.

Para ello se añade a los datos una columna denominada **Tiempo entre Auto Apertura y Decisión de Fondo** en donde se calcula el tiempo trascendido entre la fecha de apertura y la fecha de ejecutoria de la decisión de fondo.

7. Procesos que, entre la fecha de ocurrencia de los hechos y el auto de apertura del mismo, tengan un valor mayor a 1825 días, son registros que serán descartados, dado que son procesos que cumplen con la terminación anormal por caducidad (Colombia, 2000), esto generaría un estudio innecesario para el algoritmo, dado que el proceso a analizar ya tendría un vicio que lo llevaría a una terminación anticipada.

8. Procesos que, entre la fecha de ocurrencia de los hechos y el auto de apertura del mismo, tengan un valor menor o igual a 0, son registros que serán descartados, dado que son procesos con información errónea, puesto que el auto de apertura es posterior a la fecha de ocurrencia de los hechos y según (Colombia, 1991) el control realizado por las contralorías es posterior y selectivo.

9. Se debe analizar los días transcurridos entre el auto de apertura y la decisión de fondo, estos días deben ser superiores a 0 pero inferiores a 1825, también en ellos se deben contemplar que existirán procesos que no salieron de esta etapa cuando fueron declarados como prescritos. Por ende, estos procesos, en estudios posteriores, deben ser tenidos en cuenta en un análisis específico, que este enfocado en poder determinar las falencias procedimentales que no permitieron llegar a una decisión en esta etapa.

10. En esta limpieza se contemplan fechas con formatos no compatibles con los cálculos de días. Entre los filtros aplicados se encuentran 31 registros con símbolos sobrantes (2010-12-23-) y dígitos adicionales (2010-012-23). Estos fueron ajustados quedando con problemas 22. Basándonos en el artículo 9 de la ley 610 en donde se menciona que los procesos de tracto sucesivo (diversos hechos en diversas fechas) se deben contar el tiempo a partir del último hecho ocurrido (Colombia, 2000), se logran ajustar 10 registros más. Finalmente, quedan 12 registros con errores de digitación o en donde la fecha de ocurrencia de los hechos no había sido determinada, esto se observa con un mensaje textual que así lo indicaba. Estos registros son excluidos para el estudio.

Por otra parte, también se excluyen 7 registros sin fecha de apertura y cuyas actuaciones procesales no fueron registradas o carecen de coherencia procesal (Colombia, 2000).

11. Se filtra la información por su columna Origen en donde se realiza una verificación y clasificación de los orígenes existentes, encontrando los siguientes. **ACES** (Actuaciones Especiales), **GRI** (Grupo de Reacción Inmediata) e **Indagación Preliminar** (Investigación previa al proceso de responsabilidad fiscal mencionado en la ley 610 de 2000), estas tienen su origen en la denuncia ciudadana (CGR, 2013) (Colombia, 2011), **Control Excepcional**, que según el decreto 403 de 2020 en su artículo 22, tiene su origen en la denuncia ciudadana, pero en cabeza de entes gubernamentales y relación con investigaciones periodísticas de alto impacto, ya bien sea por su relevancia a nivel de cuantía o grado de afectación estatal (MinJD. Decreto, 2020), **De oficio y traslado de otras entidades**, según lo reglamentado por cada contraloría en su normatividad interna y según su competencia para conocer de una investigación fiscal, las entidades pueden dar apertura o recibir investigaciones, ya bien sea al interior de la organización o provenientes de otras organizaciones gubernamentales (Colombia, 2011). Estos dos orígenes serán relacionados con la denuncia ciudadana, según lo discutido con la funcionaria entrevistada. (Salgado., 2021). Finalmente se encuentran los orígenes **Proceso Auditor y Revisión cuentas**, estos están directamente relacionados con el PAI (Plan de Auditoria Institucional) y serán clasificados como en un único ítem nominado proceso Auditor (Salgado., 2021). En conclusión, se sacarán los Ítems como origen de datos que serán Denuncia ciudadana y Proceso Auditor.

12. Se crea una columna denominada **Amparo Póliza de Seguro** en donde teniendo en cuenta el valor diligenciado en la columna **Valor amparado por pólizas de seguros**, se coloca un valor de 0 o 1 para determinar si hay un valor registrado en la columna originaria.

13. Se crea una columna denominada **Decisión Re-Clasificada** en donde teniendo en cuenta los valores de la columna Decisión de Apertura se clasifica en 3 los registros existentes. **1. Imputación, 2. Archivo y 0. Tramite** decisorio. En el análisis de la información se encontró que todos los registros que tenían una decisión de imputación podían ser identificados plenamente por la etiqueta I1-Imputación. Por su parte los registros con archivo, estaban divididos en, I2-Archivo por no Merito, I5: Archivo por prescripción, I4: Archivo por caducidad, I3: Archivo por pago, Cesación de la acción fiscal, al consultar la ley 610 (Colombia, 2000) y el Estatuto anticorrupción (Colombia, 2011), se confirma que todas estas etiquetas hacen referencia a los procesos de responsabilidad fiscal que pertenecen a una decisión de archivo. Finalmente se encontró la etiqueta en trámite y sin ninguna etiqueta, esto nos llevó a consultar el manual de SIA- SIREL (AGR, 2019) y confirmar con la funcionaria que diligenciaba la información (Salgado., 2021), con los dos pudimos constatar que los campos no diligenciados, se debía a que en el trámite del proceso, no habían alcanzado esta actuación procesal, lo que se traduce en pocas palabras a que el proceso se encuentra en trámite. Por ello se definió la etiqueta **0 Tramite**.

14. Se crea la columna **Días entre Auto Apertura y Decisión de Fondo**, como su nombre lo indica en ella se calculan los días transcurridos entre estas dos fechas.

15. Se crea la columna **Decreto Medida Cautelar**, en ella se toma un valor 1 o 0, este valor depende de si el campo originario denominado **Fecha de decreto de la medida cautelar**, contiene una fecha valida que permita establecer que sobre el proceso se realizaron medidas cautelares (Colombia, 2000).

16. Se crea la columna **Días entre Decisión de Fondo y Fallo**, en esta se registrarán los días transcurridos entre estas dos fechas.

17. Se crea la columna **Días entre Fallo y envió a grado de consulta y/o Días entre Decisión de apertura y el envió a grado de consulta**, teniendo en cuenta que no todos los procesos transitan por el fallo ni que todos los procesos van al archivo, esta columna se encarga de discriminar estos tránsitos en los procesos.

18. Se crea una columna denominada **Sumatoria total de días transcurridos desde el auto de apertura**, la función de esta columna será la de contener la sumatoria de todos los tiempos procesales por los que transito un proceso que aún no ha terminado. Con ella mediremos los procesos que aún están tránsito y no tienen una ejecutoria.

19. Para finalizar colocamos una columna llamada **clasificador**, a ella le asignamos 3 tipos de etiquetas, **Prescripción** (Procesos marcados como prescritos), **Tramite** (Procesos que aún se encuentran activos a la fecha de rendición del registro), **Terminación Normal** (procesos que ya finalizaron en los tiempos esperados) y **Terminación en Extra-tiempo** (Procesos que debieron haber prescrito, pero están marcados con terminación normal).

Estas reglas y limpieza de datos nos permiten obtener una data con 91.161 registros válidos para nuestro estudio, esto equivale al 63.05% de la información válida para realizar un proceso de aprendizaje de máquina. Finalmente se quitan los duplicados quedando para el procesamiento 81933 registros.

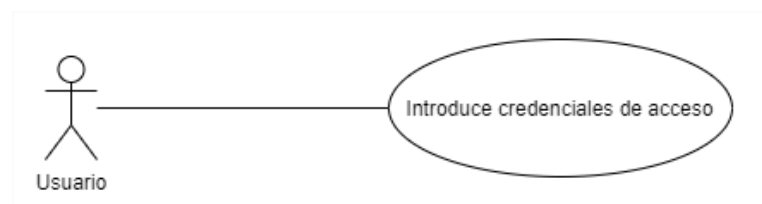
En el ítem de resultados se procederá a mostrar los resultados obtenidos.

9.4 Casos de Uso

Se pueden establecer los siguientes casos de uso en el desarrollo del sistema de vigilancia:

Figura 19

Caso de uso acceder al sistema

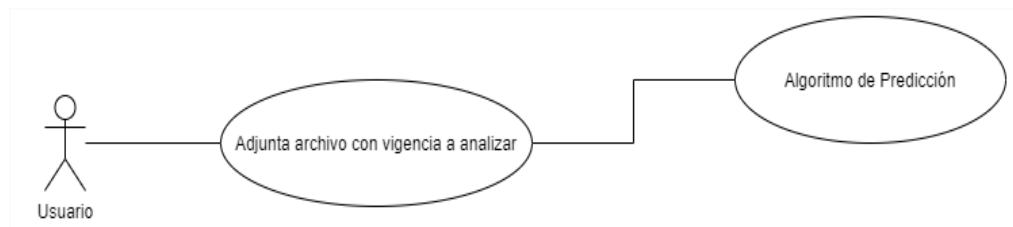


Fuente: Elaboración propia (2021)

Nombre del caso de uso: Acceder al sistema	
Precondiciones: Usuario debe estar registrado en el sistema	
Acciones del actor	Acciones del sistema
1. El usuario digita las credenciales y da clic en ingresar	
	2. El sistema valida contra la base de datos las credenciales digitadas
	2.1 Si la respuesta de la validación es correcta lo conduce a la página de donde se adjunta el archivo Excel.
	2.2 Si la respuesta de la validación es incorrecta informa del error, sea por credenciales incorrectas o porque no se tiene cuenta en el sistema.

Figura 20

Caso de uso adjuntar Excel con vigencia a analizar

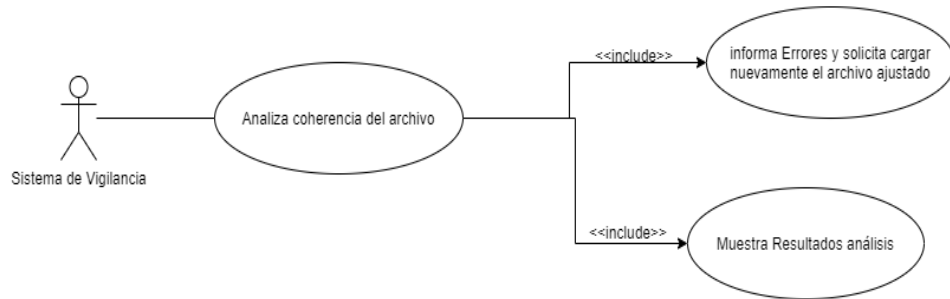


Fuente: Elaboración propia (2021)

Nombre del caso de uso: Adjuntar archivo extensión csv con vigencia a analizar	
Precondiciones: El usuario debe tener un archivo con una estructura en formato .csv	
Acciones del actor	Acciones del sistema
1. El usuario da clic en adjuntar y espera el reporte de errores de la validación del archivo o el reporte de procesos con la valoración de riesgo.	
	2. Se valida que el archivo tenga la extensión válida para comenzar a ser analizado
	2.1 Si el archivo tiene una extensión valida continua con el módulo de sistema de vigilancia.
	2.1.1 Si la validación es correcta procede analizar la coherencia del archivo

Figura 21

Caso de Uso Analizar la coherencia del archivo



Fuente: Elaboración propia (2021)

Nombre del caso de uso: Analizar la coherencia del archivo	
Precondiciones: El archivo cargado tiene una extensión válida	
Acciones del actor	Acciones del sistema
1. El usuario da clic en adjuntar y espera el reporte de errores de la validación del archivo o el reporte de procesos con la valoración de riesgo.	
	2. Evalúa el archivo cargado
	2.1 Si el archivo no tiene una extensión válida, informa el error al usuario y queda a la espera de que se adjunte un archivo correcto

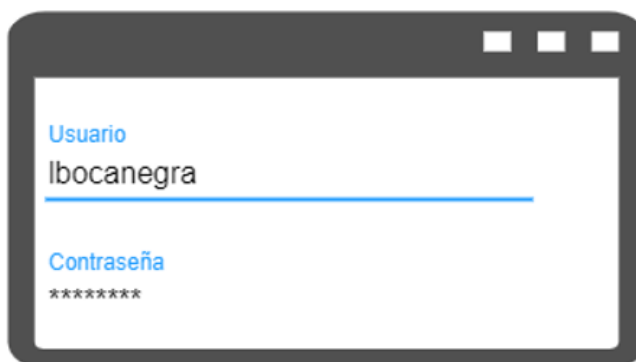
	2.2 Como el archivo tiene una extensión valida, procede a realizar las validaciones de estructura y numero de columnas
	2.2.1 Si la validación es correcta procede a validar el tipo de dato introducido en cada columna, registro a registro.
	2.2.2 Valida los datos analizados contra el modelo supervisado y genera el reporte con las estadísticas y errores.

9.5 Diseño de Pantallas

La primera pantalla a visualizar al momento de desplegar el prototipo de sistema de vigilancia, será la pantalla de seguridad.

Figura 22

Pantalla de ingreso



Fuente: Elaboración propia (2021)

Dado que el enfoque del sistema está concebido, como una herramienta de predicción interna alimentada por datos externos que no tendrán un almacenamiento de consulta en las pantallas de la herramienta, se procede a concebir la seguridad mediante la implementación de contraseña cifrada y un usuario de acceso al sistema registrado en la base de datos del mismo.

La segunda pantalla con la que interactúa el usuario es la pantalla de cargue del archivo Excel que contiene la información de la vigencia procesal a ser evaluada por parte del modelo predictivo.

Figura 23

Pantalla cargue del archivo



Fuente: Elaboración propia (2021)

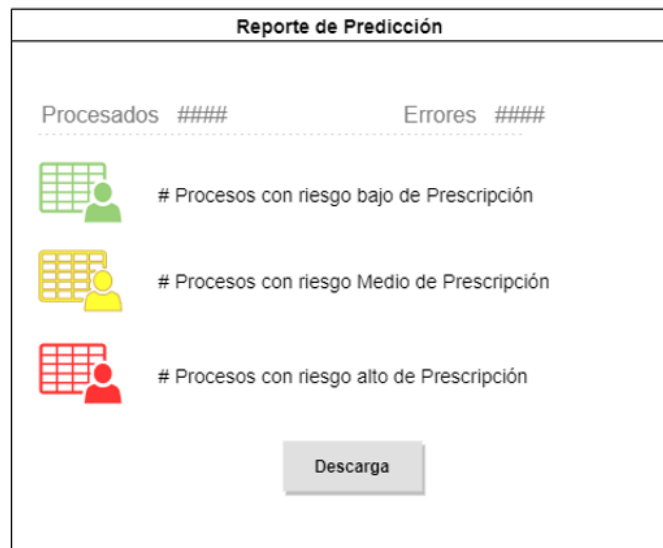
En las validaciones que se encuentran en el mismo observamos el tipo de archivo, el número de columnas y los encabezados. Esta pantalla realizará una

verificación inicial que valide las condiciones mínimas iniciales con las cuales se recibirá el archivo a pasar al modelo de predicción.

Finalmente, la pantalla que se encarga de remitir la información al modelo de predicción, también se encargara de hacer una validación profunda y campo a campo de cada uno de los registros introducidos en el archivo de cargue. En esta verificación el sistema de vigilancia tomara un control cronológico de las fechas de las actuaciones procesales, clasificando el registro en uno de los tres grupos de riesgo.

Figura 24

Pantalla de Reporte de Predicción



Fuente: Elaboración propia (2021)

De igual manera esta pantalla jugara un papel global en la clasificación del registro no entrara en especificaciones. Estas estarán en un archivo de descarga Excel que contendrá una marcación de cada uno de los registros en colores, para que el

usuario tome los correctivos pertinentes en los procesos con riesgo medio o alto de prescripción.

10 RESULTADOS

Al realizar las correspondientes pruebas del algoritmo de predicción en la herramienta Rapid Miner (Rapidminer, 2022) para poder establecer cuál era el algoritmo más adecuado a implementar en el desarrollo del sistema de vigilancia, se realizan los siguientes experimentos:

Se procede a tomar una pequeña muestra balanceada de 100 registros. En donde el 50% de ellos son tomados de procesos de prescripción y el 50 % restante son procesos en trámite, terminados en tiempo normal, es decir terminados a más tardar en 5 años y procesos con terminación en extra tiempo, es decir, procesos que tienen una marca de finalización, pero que su tiempo excede de los 5 años de terminación de procesos con una terminación normal.

Figura 25

Columnas tomadas para los experimentos

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Año	ID_ZONA	ID_COD_DANE	ID_Origen	Dias entre Ocurrenca de los hechos y Auto Apertura	Amparo Poliza de Seguros	ID_Decisión Re Clasificador	Dias entre Auto Apertura y Decisión de Fondo	Decreto Medida Cautelar	Dias entre Decisión de Fondo y Fallo	Dias entre Fallo y envío a grado de consulta/o	Dias entre envío a grado de consulta y la decisión de grado de consulta	Tiempo entre Auto Apertura y Decisión de apertura	Sumatoria total de días transcurridos desde el auto de	Clasificador
2	2012	1	250	2	1143	1	0	2034	0	0	17	14	2079	2065	Prescription
3	2012	1	250	2	1151	1	0	2034	0	0	15	16	2084	2065	Prescription
4	2012	1	250	2	1813	0	0	2197	0	0	13	21	2240	2231	Prescription
5	2012	1	250	2	643	1	0	0	0	0	0	53	2092	53	Prescription
6	2012	1	250	2	1054	1	0	1972	0	0	15	14	2005	2001	Prescription

Fuente: Elaboración propia (2021)

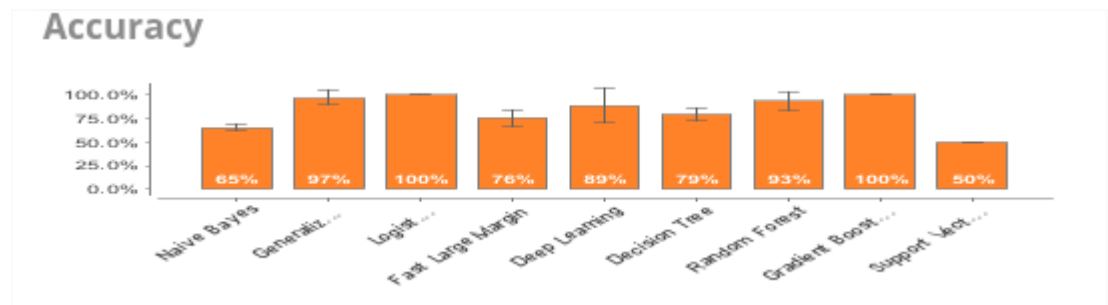
La siguiente imagen es una muestra de las columnas utilizadas para este primer experimento.

Es importante mencionar que se tuvo en cuenta que los datos tomados provinieran de diferente año, zonas, ciudades y orígenes. Esto con el fin de darle una diversidad adecuada de los datos al algoritmo y evitar sesgos que variaran la predicción.

En la siguiente figura se muestran los resultados obtenidos.

Figura 26

Grafica de exactitud en el modelo 1



Fuente: Elaboración propia (2021)

Como se puede observar los algoritmos de Regresión Logística y Gradient Boosted Tree muestran una exactitud del 100 % en la clasificación de los datos entre los procesos prescritos y los que se mencionan en el clasificador sin prescripción (Terminación en extra tiempo, Terminación Normal o en tránsito). No menos importante observamos en segundo y tercer lugar a los algoritmos Modelo lineal generalizado y a Random Forest con una precisión que supera el 90%.

Una muestra importante de la matriz de confusión de los dos primeros algoritmos de clasificación del 100% de puede ver en las siguientes figuras:

Figura 27

Matrix de confusión Modelo de regresión logística.

Confusion Matrix			
	true Prescripcion	true No prescripcion	class precision
pred. Prescripcion	14	0	100.00%
pred. No prescripcion	0	15	100.00%
class recall	100.00%	100.00%	

Fuente: Elaboración propia - Generado con RapidMiner (2021)

En un modelo equilibrado en el que la mitad de datos corresponden a prescripción y la otra mitad a procesos en trámite o con una terminación normal o en extra tiempo, los modelos Regresión Logística y Gradient Boosted Tree, tienen una precisión óptima que permite una exactitud en la predicción.

Figura 28

Matrix de confusión Modelo de Gradient Boosted Tree.

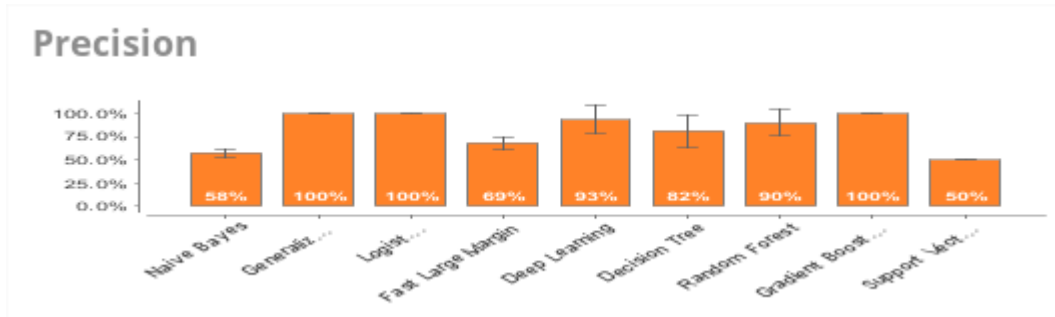
Confusion Matrix			
	true Prescripcion	true No prescripcion	class precision
pred. Prescripcion	15	0	100.00%
pred. No prescripcion	0	14	100.00%
class recall	100.00%	100.00%	

Fuente: Elaboración propia - Generado con RapidMiner (2021)

Si se validan otros datos como precisión y la sensibilidad, se encuentran valores del 100%, así como lo muestran las siguientes graficas.

Figura 29

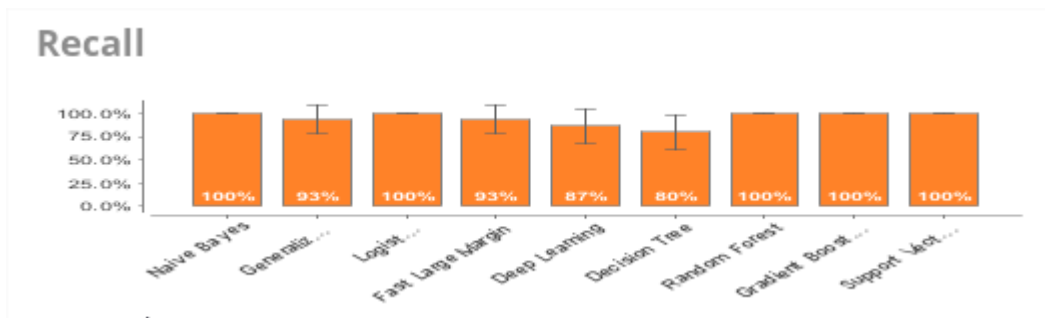
Grafica de precisión con el modelo 1



Fuente: Elaboración propia - Generado con RapidMiner (2021)

Figura 30

Grafica de sensibilidad con el modelo 1.



Fuente: Elaboración propia - Generado con RapidMiner (2021)

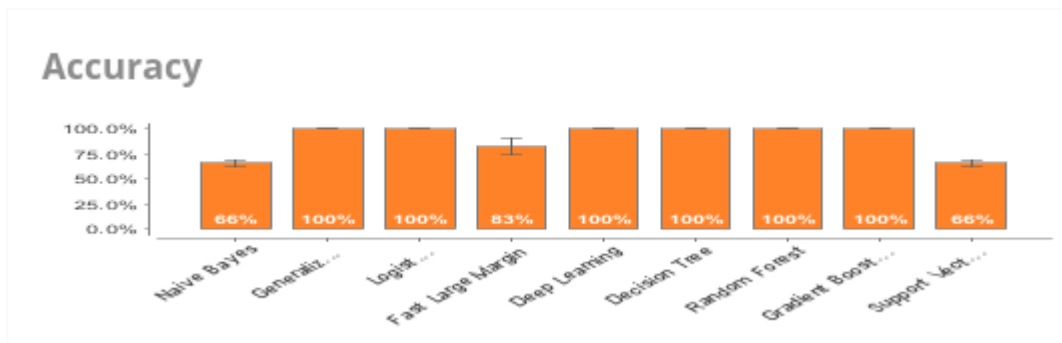
Dados estos resultados, se procede a realizar un segundo modelo, que nos permita ver el comportamiento de los algoritmos destacados, cuando se aumenta el número de registros para el modelo. Por lo que, para este segundo experimento, procedemos a disminuir el porcentaje de registros con prescripción y colocamos solo registros cuyo clasificador sean procesos terminados en extra tiempo. Con la etiqueta

de No Prescrito. Esto nos dará una visión tendencia más clara de la función de los algoritmos que están evaluando. Para este experimento duplicaremos los registros por prescripción y multiplicaremos cuatro veces los registros con terminación de extra tiempo que es uno de los parámetros que más podrían confundirse a la hora de detectar procesos con prescripción.

En la siguiente grafica se observa los resultados obtenidos de este segundo experimento en lo que respecta a la exactitud.

Figura 31

Grafica de Exactitud con el modelo 2

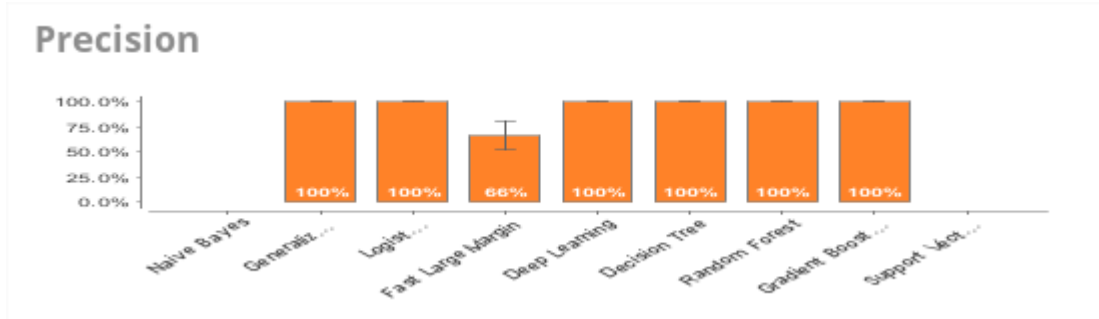


Fuente: Elaboración propia - Generado con RapidMiner (2021)

Como se puede observar el aumento de registros con extra tiempo, tiene afectaciones positivas en la exactitud de los procesos con prescripción, solo el algoritmo de Naive Bayes tuvo desmejoras en la predicción. Casos similares de mejora, observamos en las gráficas relacionadas con los factores de precisión y sensibilidad.

Figura 32

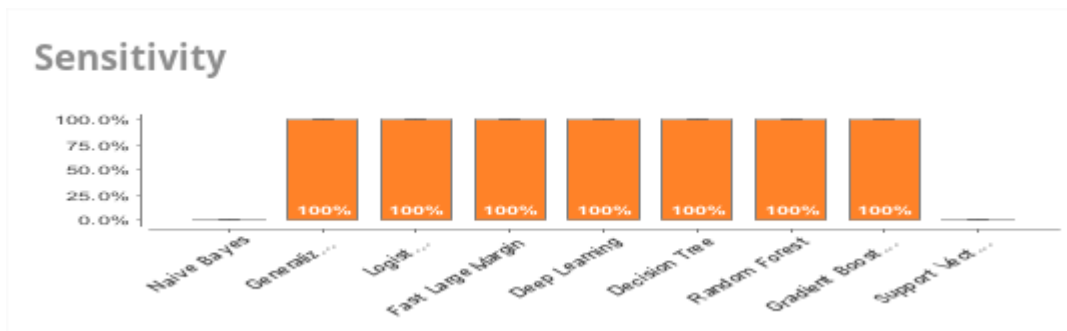
Grafica de Precisión con el modelo 2



Fuente: Elaboración propia - Generado con RapidMiner (2021)

Figura 33

Grafica de Sensibilidad con el modelo 2.



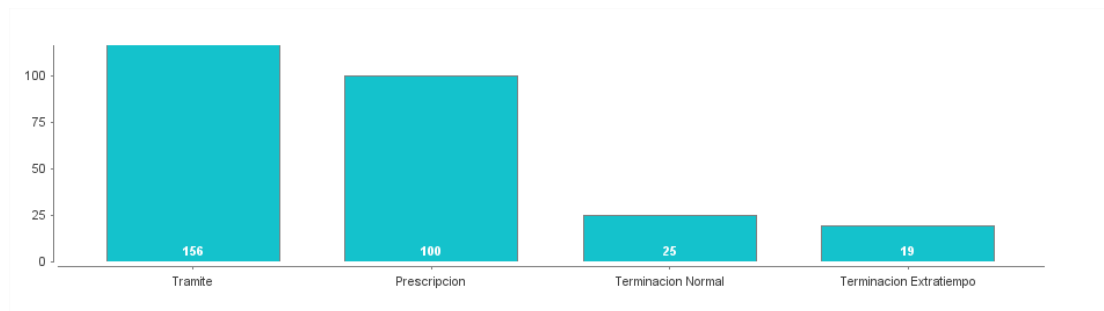
Fuente: Elaboración propia - Generado con RapidMiner (2021)

Se sugiere un tercer experimento que permita disminuir las variables de entrada del modelo de predicción, con el fin de mejorar el rendimiento del algoritmo sin afectar el nivel de exactitud, sensibilidad y precisión del mismo.

Por ello se selecciona un nuevo set de datos con 300 registros variados entre, procesos en tránsito, prescritos, con terminación normal y terminación en extra tiempo. Disminuyendo el número de variables de entrada. En cuanto a los valores de clasificación los establecimos inicialmente como se muestra en la siguiente grafica.

Figura 34

Grafica de clasificador tomado para el modelo 3



Fuente: Elaboración propia - Generado con RapidMiner (2021)

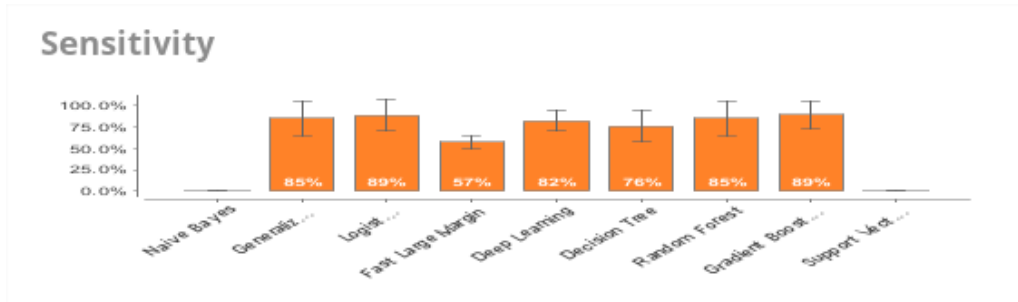
Se descartarán para la generación de este modelo las siguientes variables de entrada:

1. Días entre la decisión de fondo y el fallo: esta variable es sugerida de ser retirada por Rapid Miner esto a raíz de que el parámetro de estabilidad es superior al 90%.
2. El año, el ID_Zona, el ID_Codigo_Dane y el ID_origen, esto porque su correlación es baja y su estabilidad relativamente alta. Los demás parámetros los mantendremos y se generara el modelo. Los resultados obtenidos son:

Una disminución visible en la sensibilidad del modelo, como se puede observar en el siguiente gráfico.

Figura 35

Grafica de sensibilidad para el modelo 3

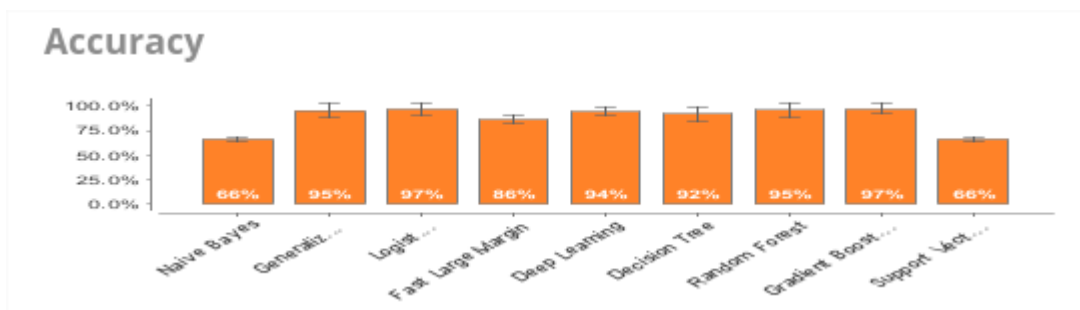


Fuente: Elaboración propia - Generado con RapidMiner (2021)

Por otro lado, también se obtiene una baja en la exactitud y aumento del algoritmo de Fast Large Margin para la presión.

Figura 36

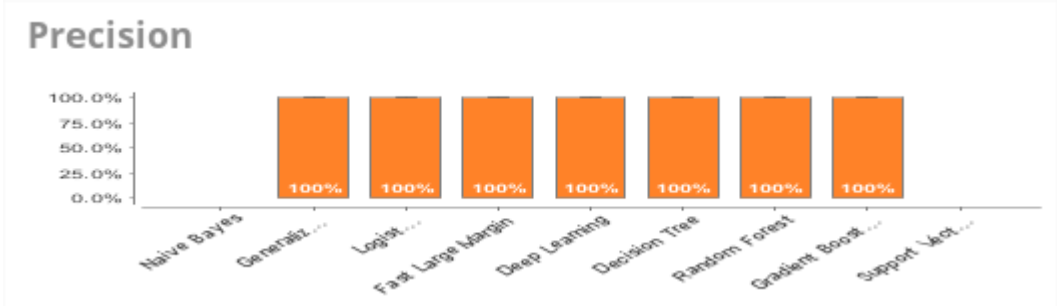
Grafica de Exactitud para el modelo 3



Fuente: Elaboración propia - Generado con RapidMiner (2021)

Figura 37

Grafica de Precisión para el modelo 3



Fuente: Elaboración propia - Generado con RapidMiner (2021)

11 DISCUSIÓN Y CONCLUSIONES

Según lo observado en el análisis con los grafos generados, se pudo constatar que al generar conjuntos de procesos prescritos y ser diagramados en Python con la ayuda de las librerías Pandas, NetworkX y Matplotlib, se pudo encontrar las variables de entrada para el diseño de los modelos de Machine Learning, mencionados en el apartado de resultados. Esto demuestra que el uso de la teoría de grafos en proyectos relacionados con detección del fraude es un aporte importante a la resolución de problemas de este estilo, dado que permite encontrar tendencias de comportamiento de los datos y las correlaciones que entre estos pueda llegar a existir. Sin embargo, la sola minería de datos, aplicada a la base de información proporcionada por la Auditoría General de la República, en sí sola, no permite una adecuada búsqueda de tendencias clave, que permitan la detección clara de los patrones de comportamiento, que tienen los procesos con terminación anormal.

Para poder encontrar dichos comportamientos se hizo necesario, proceder a conocer el desarrollo de los procesos de responsabilidad fiscal en las contralorías territoriales. Esto con el fin de poder realizar mediciones de tiempo entre las actuaciones procesales y los términos establecidos por la legislación colombiana.

Podemos concluir que el solo uso de la teoría de grafos, no es suficiente para poder llegar a predecir la problemática de las terminaciones anormales en los procesos de responsabilidad fiscal, por lo que se requiere para estudios posteriores, obtener mucha más información, relacionada con el desarrollo de dichos procesos y la

concordancia de este desarrollo con la información capturada por la Auditoría General de la República. Dicha información debería ser relacionada con la información encontrada en los juzgados, para determinar la duración de las actuaciones procesales que se llevan a cabo en el poder judicial, información a la que para este trabajo no tuvimos acceso, debido a la reserva procesal mencionada en el artículo 20 de la ley 610 de 2000 (Colombia, 2000).

Por otra parte el proceso de minería de datos realizado, luego de descubrir mediante los grafos, las tendencias de los procesos con terminación anormal, ayudó de manera significativa a obtener la información suficiente, para que los algoritmos de Machine Learning alcanzaran a identificar, que procesos tienen una clasificación de prescripción y que procesos, a pesar de que según el tiempo transcurrido y las variables de entrada prejuiciosamente así lo indicaran, no cumplen con una clasificación de prescripción.

Estos resultados muestran que mezclando diversas técnicas de inteligencia artificial, como lo son la Minería de Datos y el Machine Learning, apalancándose con la teoría de grafos, pueden llegar a ser útiles en la detección de focos de corrupción en ámbitos judiciales y de control fiscal. Estos estudios pueden arrojar resultados de calidad, siempre y cuando la información recolectada para la generación de los mismos, también sea de calidad y no implique la exclusión inicial de registros, por errores en la recolección y procesamiento de la información.

Para futuras investigaciones, se recomienda, no solo enfocar el trabajo a la búsqueda de causales de terminación anormal por prescripción. Esto dado que existen otras causales, como son la caducidad, la nulidad y el pago por no merito, que, modelados con algoritmos de inteligencia artificial, ayudarían a detectar procesos que desde la apertura pueden estar a un tiempo muy corto para poder culminar todas las actuaciones procesales. De igual manera se anima a que grupos de investigación se enfoquen en el estudio de otros procesos de rendición de cuentas, sean modelados con análisis de teoría de grafos, Minería de Datos y Machine Learning, esto dado a que la combinación de diversas técnicas, puede aportar a la detección temprana del fraude en entidades del orden nacional como, la fiscalía y la procuraduría, al igual que entes territoriales como las personerías, gobernaciones y alcaldías, estas últimas que manejan un control de veeduría al interior de sus organizaciones.

REFERENCIAS

- [1] Calabuig, J. M., Falciani, H., Ferrer-Sapena, A., García-Raffi, L. M., Raso, E., Sánchez Del Toro, I., Sánchez-Pérez, E. A., & Madφ, G. (2018). Detección y prevención de las malas prácticas y la corrupción desde la perspectiva de las matemáticas (*).
- [2] Solé, J. P. (2018). La prevención de riesgos de mala administración y corrupción, la inteligencia artificial y el derecho a una buena administración (*). http://www.leru.org/files/publications/Interdisciplinarity_and_the_21st_century_research-intensive_university.pdf
- [3] Microsoft Docs. (2020) Algoritmos de minería de datos (Analysis Services, minería de datos). Retrieved May 8, 2020, from <https://docs.microsoft.com/es-es/analysis-services/data-mining/data-mining-algorithms-analysis-services-data-mining?view=asallproducts-allversions>
- [4] IBM. (2012). Manual CRISP-DM de IBM SPSS Modeler, from <ftp://ftp.software.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>
- [5] Emerio Villamil, J. (2010). Corruption in Colombia, conceptual and methodological approaches to address it.
- [6] Blanes Climent, M.A. (2017). El sistema de alertas desde enfoques diversos. el SÍNDIC DE GREUGES (*). Revista internacional de transparencia e integridad.

Disponible en: https://revistainternacionaltransparencia.org/wp-content/uploads/2017/12/miguel_angel_blanes.pdf

[7] Zuleta, A. P. (n.d.) (2015). La corrupción su historia y sus consecuencias en Colombia.

[8] Solé, J. P. (2018). La prevención de riesgos de mala administración y corrupción, la inteligencia artificial y el derecho a una buena administración (*). Revista internacional de transparencia e integridad. Disponible en: http://www.leru.org/files/publications/Interdisciplinarity_and_the_21st_century_research-intensive_university.pdf

[9] Amoedo Barreiro, D. (2018). Los sistemas informáticos de detección de malas prácticas, herramientas esenciales para la prevención de la corrupción (*). Revista internacional de transparencia e integridad. Disponible en: https://revistainternacionaltransparencia.org/wp-content/uploads/2017/12/alberto_mario_gonzalez.pdf

[10] Comisión Económica para América Latina y el Caribe (CEPAL) (2020). Recuperar la confianza de los ciudadanos en las instituciones públicas es fundamental para retomar una senda de crecimiento inclusivo y de mayor bienestar para todos en América Latina y el Caribe | Comunicado de prensa. Retrieved October 15, 2020, from <https://www.cepal.org/es/comunicados/recuperar-la-confianza-ciudadanos-instituciones-publicas-es-fundamental-retomar-senda>

- [11] Jaimes, O. et al. (2016) La corrupción y el fenómeno Nule. Corruption and Nule Phenomenon. Disponible en: <http://hdl.handle.net/11232/736>
- [12] Peña Álvarez, C. C., Pinilla, C., & Bello, M. (2017). Bases de datos orientadas a grafos. Tecnología Investigación y Academia, 5(2), 153–160. Recuperado a partir de <https://revistas.udistrital.edu.co/index.php/tia/article/view/8769>
- [13] Cerillo i Martínez, A. (2018). Diez propuestas para la colaboración ciudadana en la alerta de malas práctica en la administración pública. Revista Internacional Transparencia ORG. Disponible en: https://revistainternacionaltransparencia.org/wp-content/uploads/2018/04/agusti_cerrillo.pdf
- [14] Torres De Durand, E. (2019). Ingovernabilidad y corrupción. Centro de Gobernanza publica y Corporativa. Disponible en: <https://governanzapr.org/wp-content/uploads/2019/09/Ingovernabilidad-y-corrupcion.pdf>
- [15] Congreso de Colombia (2011). Ley 1474 de 2011. 12 de julio 2011(Colombia).
- [16] Congreso de Colombia (2000). Ley 610 de 2000. 12 de julio 2011(Colombia).
- [17] Ministerio de Justicia y del Derecho. Decreto (2020) – Ley 403 de 2020. 16 de marzo 2020.
- [18] Constitución Política de Colombia, [Const] (1991). Art. 29 y Art 209. 7 de julio de 1991 (Colombia).

- [19] Bellomarini, L., et al. (2020). Knowledge graphs: the layered perspective. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 12072 LNCS (pp. 20–34). Springer. https://doi.org/10.1007/978-3-030-53199-7_2
- [20] Janev, V. (2020). Ecosystem of big data. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 12072 LNCS (pp. 3–19). Springer. https://doi.org/10.1007/978-3-030-53199-7_1
- [21] Palestina, I.E (2018). Corrupción y crecimiento económico: las percepciones de la sociedad colombiana. Revista CES Derecho, (9), 1, enero – junio 2018, 59-72. Disponible en: <http://www.scielo.org.co/pdf/cesd/v9n1/2145-7719-cesd-9-01-59.pdf>
- [22] Isaza E., C. (2011). El fracaso de la lucha anticorrupción en Colombia. OPERA. 11, 11 (nov. 2011), 221–239. Disponible en: <https://revistas.uexternado.edu.co/index.php/opera/article/view/3571>
- [23] Constitución Política de Colombia (1991), [Const]. Art. 272. 7 de julio de 1991 (Colombia).
- [24] Ordóñez Vásquez, T. et al. (2019). Fiscal findings and processes of fiscal responsibility in Colombia 2012-2017. Revista Republicana, 2019(27), 211–233. Disponible en: <https://doi.org/10.21017/REV.REPUB.2019.V27.A74>

- [25] Lima, M. S. M., & Delen, D. (2020). Predicting and explaining corruption across countries: A machine learning approach. *Government Information Quarterly*, 37(1), 101407. <https://doi.org/10.1016/j.giq.2019.101407>
- [26] Benbouzid, B. (2019). Values and consequences in predictive machine evaluation. A sociology of predictive policing¹. *Science and Technology Studies*, 32(4), 119–136. <https://doi.org/10.23987/sts.66156>
- [27] Gómez, C. et al. (2020). Economía de datos e inteligencia artificial en América Latina. Oportunidades y riesgos para un aprovechamiento responsable. Buenos Aires: CETyS Universidad de San Andrés. Disponible en: https://7da2ca8d-b80d-4593-a0ab-5272e2b9c6c5.filesusr.com/ugd/7be025_59feb42ee7854fb8b9009195993da055.pdf
- [28] Congreso de Colombia (1992). Código de procedimiento administrativo, Artículo 44 y 45 (Colombia).
- [29] Entrevista Telefónica con la Exfuncionaria de la Contraloría de Bogotá y Auditoría General de la Republica – Ana Iddaly Salgado. Marzo de 2021.
- [30] Urueta, Carlos., et al. (2021). Aplicaciones de la teoría de grafos en la escogencia del punto óptimo para la implementación de una planta de beneficio a partir de la palma de aceite. Retrieved March 22, 2021, from <http://www.mat.ucm.es/catedramdeguzman/modelizaciones/proyectos/proyecto10/index.htm>

- [31] Atzeni, P., et al. (2020). Weaving enterprise knowledge graphs: The case of company ownership graphs*. *Advances in Database Technology - EDBT*, 2020-March, 555–566. <https://doi.org/10.5441/002/edbt.2020.66>
- [32] Sposito, O., et al (2021). CACIC 2020 - SMOTE, Algoritmo para balanceo de clases en un estudio aplicado a la ganadería. Organizado por el Dto. Ingeniería e Investigaciones Tecnológicas UNLaM y la RedUNCI. Retrieved October 3, 2021, from www.cs.waikato.ac.nz/~ml/weka/
- [33] Asniar, Nur U. et al. (2021), SMOTE-LOF for noise identification in imbalanced data classification, *Journal of King Saud University – Computer and Information Sciences*, <https://doi.org/10.1016/j.jksuci.2021.01.014>
- [34] Arrieta, J., & Mera, Carlos. (2015). Métodos de clasificación con etiquetado débil para la inspección visual automática en problemas industriales View project Automatic visual inspection through Multiple Instance Learning View project Carlos Mera Instituto Tecnológico Metropolitano. <https://www.researchgate.net/publication/283642919>
- [35] Moreno, J & Rodríguez, et al. (2009). SMOTE-I: mejora del algoritmo SMOTE para balanceo de clases minoritarias. Disponible en: https://www.researchgate.net/publication/229045207_SMOTE-I_mejora_del_algoritmo_SMOTE_para_balanceo_de_clases_minoritarias/citation/download

- [36] Vásquez, M. et al. (2019). Balanceo de datos del Síndrome de Guillain-Barré utilizando SMOTE para la clasificación de subtipos. *Research in Computing Science*. 148. 113-125. 10.13053/rcs-148-7-9. Disponible en: https://rcs.cic.ipn.mx/2019_148_7/Balanceo%20de%20datos%20del%20Sindrome%20de%20Guillain-Barre%20utilizando%20SMOTE%20para%20la%20clasificacion%20de%20subtipos.pdf
- [37] Bechberger D. and Perryman J. (2020), “Graph Databases in Action”, 1st ed. Manning Publications.
- [38] Ralph P. Grimaldi (1998), *Matemáticas discretas y combinatoria: una introducción con aplicaciones*, Pearson Educación. Disponible en <https://books.google.es/books?id=IHqqjoR0b1YC>
- [39] Moreno, W. et al. (2012). El triángulo del fraude. In *Forum Empresarial* (Vol. 17, No. 1, pp. 65-81). Centro de Investigaciones Comerciales e Iniciativas Académicas. Disponible en: <https://www.redalyc.org/pdf/631/63124039003.pdf>
- [40] Abarca Aguirre, J. (2010). *La corrupción burocrática: corruptos, corruptores, delitos y justicia en Chile colonial (1621-1700)*. Tesis Doctoral- Universidad de Chile, Departamento de Ciencias Históricas. Santiago de Chile. Disponible en: https://repositorio.uchile.cl/bitstream/handle/2250/108601/fi-abarca_j.pdf?sequence=3&isAllowed=y

- [41] Fundación Vortex para la Procuraduría General de la Nación, Colombia (2020).
Convenio de Cooperación para Diseñar una metodología que permita identificar
redes y patrones de macro-corrupción en Colombia. Agosto 4 de 2020. Bogotá,
Colombia. Disponible en:
<https://www.procuraduria.gov.co/portal/media/file/Protocolo%20metodo%20B3gico%20de%20An%C3%A1lisis%20de%20Redes%20Criminales%20en%20la%20Procuradur%C3%ADa%20General%20de%20la%20Naci%C3%B3n.pdf>
- [42] Sánchez Durán, J. (2020). Bigdata y gestión pública colombiana: Caso Contraloría
General de la República. Universidad de los Andes. Mayo de 2020. Bogotá,
Colombia. Disponible en:
<https://repositorio.uniandes.edu.co/bitstream/handle/1992/48418/u833453.pdf>
- [43] Contraloría General de la Republica. (2019). Océano. Año 33. Septiembre -
noviembre de 2019. Central de información contractual. Revista Economía
Colombiana. N. 356. Disponible en:
<https://www.economicolombiana.co/revista/oceano-393>
- [44] Contraloría General de la Republica. (2013). Resolución orgánica 7130 de 2013.
Mayo 16 de 2013. Disponible en:
https://normativa.colpensiones.gov.co/colpens/docs/resolucion_contraloria_7130_2013.htm

- [45] Contraloría General de la Republica. (2021). Definiciones de Responsabilidad Fiscal. Disponible en: <https://www.contraloria.gov.co/control-fiscal/responsabilidad-fiscal> Consultado el 10 de diciembre de 2021.
- [46] TodoBI Business Intelligence (2019). Análisis de los Panamá Papers con Neo4J - Big Data (Comercial). Disponible en: <https://todobi.com/analisis-de-los-panama-papers-con-neo4j/> Consultado el 27 de diciembre de 2021.
- [47] Departamento Administrativo Nacional de Estadística. (2021). Departamentos y municipios de Colombia. Actualizado el 05 de noviembre de 2021. Disponible en: <https://www.datos.gov.co/Mapas-Nacionales/Departamentos-y-municipios-de-Colombia/xdk5-pm3f> .Consultado el: 27 de diciembre de 2021.
- [48] Auditoria General de la Republica. (2019). SIA Misional AGR Manual del módulo SIREL. Disponible en: http://misional.auditoria.gov.co/modules/rec/documentos/Manual_REC_AGR.pdf
- [49] Daza Vergaray, A. (2016). Data Mining. Minería de Datos. Editorial Macro, Primera edición. Julio de 2016.
- [50] Sandoval, L. (2018). Algoritmos de aprendizaje automático para análisis y predicción de datos. Repositorio Digital de la Ciencia y Cultura de El Salvador REDICCES. Disponible en: <https://core.ac.uk/reader/162564840>

[51] Rapidminer. Documentación de Rapidminer (2014). Disponible en:
<https://docs.rapidminer.com/downloads/RapidMiner-v6-user-manual.pdf>

Consultado el 02/01/2022

Anexos

ANEXO 1

En la siguiente tabla se observa el análisis previo del preprocesamiento de la información suministrada por la Auditoría General de la República, en esta se puede observar el número de expedientes por contraloría.

Contraloría	Número de expedientes
Amazonas	44
Antioquia	10467
Arauca	409
Armenia	367
Atlántico	597
Barrancabermeja	695
Barranquilla	223
Bello	671
Bogotá D.C.	1681
Bolívar	1718
Boyacá	2748
Bucaramanga	600
Buenaventura	2258
Caldas	1763
Cali	545
Caquetá	2552
Cartagena de Indias	1887
Casanare	4561
Cauca	1220
Cesar	1469
Chocó	791
Córdoba	1686
Cúcuta	162
Cundinamarca	12409
Dosquebradas	154
Envigado	142
Floridablanca	322
General de la República	45846

Contraloría	Número de expedientes
Girón	10
Guainía	230
Guajira	1164
Guaviare	385
Huila	1478
Ibagué	1047
Itagiú	146
Magdalena	1725
Manizales	675
Medellín	1116
Meta	2905
Montería	99
Nariño	7452
Neiva	482
Norte de Santander	4812
Palmira	476
Pasto	366
Pereira	534
Popayán	547
Putumayo	2033
Quindío	367
Risaralda	856
San Andrés, Providencia y Santa Catalina	346
Santa Marta	192
Santander	4232
Sincelejo	15
Soacha	624

Contraloría	Número de expedientes
Soledad - Atlántico	245
Sucre	1555
Tolima	2439
Tuluá	277
Tunja	539
Valle del Cauca	3345
Valledupar	731

Contraloría	Número de expedientes
Vaupés	525
Vichada	987
Villavicencio	1443
Yumbo - Valle	176

Fuente: Elaboración propia basada en los datos suministrados por la Auditoría General de la República.

ANEXO 2

EL DECANO DE LA FACULTAD DE INGENIERÍA Y LA COORDINADORA DEL PROGRAMA ACADÉMICO DE INGENIERÍA DE SISTEMAS

HACEN CONSTAR

Que, **Luis Ernesto Bocanegra Rairán** y **José Manuel Chauta Torres** participaron en la modalidad Póster con el trabajo titulado: “**Aplicación de la Teoría de Grafos en el Manejo Procesal, como una Herramienta de Apoyo en los Procesos de Responsabilidad Fiscal en Colombia**” en el “**3er Congreso Internacional de TIC para la Amazonía – Retos para la Transformación Digital**”, llevado a cabo en formato virtual, los días 17, 18 y 19 de noviembre del 2021. Este evento fue organizado por: La Universidad de la Amazonia, La Universidad de Cundinamarca, La Fundación Universitaria de la Rioja - UNIR Colombia y WONCY – LATAM Women in Cybersecurity.

La presente se expide en la ciudad de Florencia, Caquetá - Colombia; a los 10 días del mes de diciembre del 2021.


JORGE ALBERTO GUZMAN MALDONADO
Decano
Facultad de Ingeniería


KELLY JOHANNA TOLEDO ARTUNDUAGA
Coordinadora
Programa Académico de Ingeniería de
Sistemas