

**Modelo para definir índices de riesgo de presuntas  
situaciones de corrupción en convocatorias de contratación  
estatal en Colombia basada en Big Data y Procesamiento del  
Lenguaje Natural**



**Ing. Julio César Luna Ortega**

**Director: PhD. Carlos Alberto Cobos Lozada  
Codirectora: PhD. Martha Eliana Mendoza Becerra**

**Universidad del Cauca  
Facultad de Ingeniería Electrónica y Telecomunicaciones  
Departamento de Sistemas  
Grupo de I+D en tecnologías de la información (GTI)  
Área de interés en Gestión de la Información  
Popayán, junio de 2022**

**Modelo para definir índices de riesgo de presuntas  
situaciones de corrupción en convocatorias de contratación  
estatal en Colombia basada en Big Data y Procesamiento del  
Lenguaje Natural**

**Ing. Julio César Luna Ortega**

**Tesis presentada a la Facultad de Ingeniería Electrónica y  
Telecomunicaciones de la Universidad del Cauca para obtener el título de**

**Magíster en Computación**

**Director: PhD. Carlos Alberto Cobos Lozada  
Codirectora: PhD. Martha Eliana Mendoza Becerra**

**Universidad del Cauca  
Facultad de Ingeniería Electrónica y Telecomunicaciones  
Departamento de Sistemas  
Grupo de I+D en tecnologías de la información (GTI)  
Área de interés en Gestión de la Información  
Popayán, junio de 2022**

## Resumen

Recientemente, se ha observado un notable crecimiento de los casos de corrupción en procesos de contratación. Esto ha motivado a que las entidades promuevan estrategias anticorrupción, la mayoría de estas estrategias son manuales y de índole social. Para poder predecir o detectar efectivamente (eficiente y eficazmente) la corrupción se hace necesario realizar procesos automáticos (basadas en estadística, inteligencia artificial y otras técnicas) para reducir y en un futuro eliminar este flagelo de la sociedad. Es de destacar que en las investigaciones enfocadas a esta problemática buscan resolver el problema de identificación de presuntas situaciones de corrupción en procesos de contratación haciendo énfasis en un delito en específico, mediante el uso, de una o dos herramientas para ello. Sin embargo, los estudios también muestran que, dada la complejidad del problema, se hace necesario contar con un conjunto (batería) de herramientas especializadas para atacar los diferentes delitos que ocurren en el proceso de contratación, dado esto, se logra entender la importancia que tiene continuar con el desarrollo de modelos que permitan solventar de manera más efectiva el problema de la identificación de presuntas situaciones de corrupción en la contratación. Además, estos trabajos han sido desarrollados en otros países y no en Colombia.

Por lo anterior, en el presente trabajo de investigación se propuso un modelo macro que permitiera detectar diferentes probables delitos o anomalías relacionadas con corrupción en los procesos de contratación pública en Colombia, para ello el modelo propuesto consta de 5 componentes principales, 1) los servicios especializados que buscan identificar situaciones específicas de probable corrupción, en específico se propusieron tres servicios dedicados a un delito en específico cada uno, la detección de similitud entre documentos, la detección de manipulación de ofertas y la detección de carteles, 2) los servicios transversales que soportan la transformación del modelo en una herramienta software, 3) servicios adicionales donde se abordan situaciones generales de probable corrupción, en específico el servicio de alerta ciudadana, 4) las relaciones explícitas entre los servicios y 5) la salida global del modelo.

El modelo propuesto, así como los servicios de lo integran fueron evaluados por un grupo de expertos, usando como base unas directrices de diseño y contenido del estándar de la nomenclatura de modelado UML, dando una calificación mínima de 4,3 sobre 5 para cada servicio expuesto. Finalmente, en la experimentación práctica dos de los servicios planteados en esta investigación fueron puestos a prueba en diversos escenarios, con los resultados arrojados por algunas de las métricas clásicas del área, se determinó el nivel de satisfacción de la calidad de la predicción obtenida por los servicios.

## Dedicatoria

*A mis padres, por todo el apoyo, amor y dedicación que han tenido conmigo, todo lo que he sido, soy y seré es por ellos y para ellos.*

## Agradecimientos

A la Dra. Martha Eliana Mendoza y al Dr. Carlos Alberto Cobos por brindarme la oportunidad de realizar mi trabajo de tesis de maestría bajo su dirección, por su dedicación, apoyo y tiempo brindado para poder culminar de la mejor manera posible este proyecto de investigación. Con un párrafo aparte para el Dr. Carlos Cobos, a quien siempre le agradeceré por todas las oportunidades, conocimientos y experiencias que me ha brindado, su guía, en mi etapa estudiantil, ha contribuido en gran parte para ser el profesional que soy hoy en día.

A todos los profesores de la Maestría en Computación de la Universidad del Cauca, por sus enseñanzas y colaboración durante mi proceso formativo en la maestría.

Al centro de desarrollo tecnológico CreaTIC por haberme brindado la posibilidad de ser beneficiario de una beca de estudios, el agradecimiento no es solo por lo económico, que, por supuesto ha sido de gran ayuda, sino también por permitirme ser parte de un programa que pretende construir un mejor país, basándose en la construcción de mejores profesionales. Con una mención especial para el Ingeniero Ricardo Zambrano por toda su ayuda brindada para poder ser beneficiario de este programa de beca.

Al instituto anticorrupción por el apoyo brindado a este proyecto, lo cual permitió en gran medida enriquecer la investigación desarrollada.

## Tabla de Contenido

1.	Introducción.....	1
1.1	Planteamiento del problema.....	1
1.2	Aportes del proyecto .....	4
1.3	Objetivos .....	5
1.3.1	Objetivo general.....	5
1.3.2	Objetivos específicos.....	5
1.4	Resultados obtenidos.....	5
1.5	Estructura de la monografía .....	6
2.	Estado del Arte.....	7
2.1	Planeación de la Revisión Sistemática de la Literatura (RSL) .....	7
2.1.1	Preguntas de investigación.....	7
2.1.2	Desarrollo del protocolo de revisión.....	8
2.2	Resultados de la RSL.....	10
2.2.1	Análisis de los grupos resultantes .....	13
2.3	Conclusiones de la RSL.....	15
3.	Modelo propuesto .....	17
3.1	Servicio de detección de similitud en propuestas técnicas.....	19
3.1.1	Datos de entrada requeridos .....	23
3.1.2	Procedimiento.....	24
3.1.3	Salida.....	26
3.2	Servicio de detección de probable manipulación de ofertas .....	28
3.2.1	Normas vigentes.....	33
3.2.2	Datos de entrada requeridos .....	34
3.2.3	Procedimiento.....	35
3.2.4	Salida.....	41
3.3	Servicio de detección del probable delito de organización de carteles y análisis del beneficiario final .....	42
3.3.1	Datos de entrada requeridos .....	50
3.3.2	Procedimiento.....	52
3.3.3	Salida.....	55
3.4	Servicio de alerta ciudadana sobre irregularidades en procesos de convocatorias de contratación pública .....	56
3.4.1	Datos de entrada requeridos .....	58

3.4.2	Procedimiento.....	59
3.4.3	Salida.....	60
3.5	Servicios transversales al modelo.....	61
3.5.1	Administración, autenticación y autorización de usuarios.....	62
3.5.2	Administración de los servicios.....	65
4.	Servicios desarrollados.....	68
4.1	Desarrollo del servicio de detección de similitud en propuestas técnicas	70
4.2	Desarrollo del servicio de detección de probable manipulación de ofertas	74
5.	Evaluación del modelo y los servicios desarrollados.....	78
5.1	Evaluación conceptual.....	78
5.1.1	Resultados del grupo focal de retroalimentación.....	80
5.1.2	Resultados del proceso de evaluación formal.....	80
5.2	Evaluación y experimentación de los servicios desarrollados.....	85
5.2.1	Evaluación del servicio de detección de similitud.....	85
5.2.2	Evaluación del servicio de detección de manipulación de ofertas.....	87
6.	Conclusiones y trabajos futuros.....	92
6.1	Conclusiones.....	92
6.2	Trabajos futuros.....	94
7.	Bibliografía.....	95

## Índice de Tablas

<b>Tabla 1.</b>	Preguntas de investigación.....	7
<b>Tabla 2.</b>	Cadena de búsqueda usada para iniciar la revisión.....	8
<b>Tabla 3.</b>	Proceso de búsqueda.....	9
<b>Tabla 4.</b>	Prueba de evaluación de la calidad.....	10
<b>Tabla 5.</b>	Estudios primarios ordenado por año y puntaje.....	10
<b>Tabla 6.</b>	Agrupamiento de los estudios.....	11
<b>Tabla 7.</b>	Tabla de precios de la propuesta económica.....	34
<b>Tabla 8.</b>	Datos de entrada necesarios para el servicio de denuncia ciudadana. ...	58
<b>Tabla 9.</b>	Preguntas realizadas en el debate.....	79
<b>Tabla 10.</b>	Protocolo de la sesión del Grupo Focal.....	79
<b>Tabla 11.</b>	Participantes de la sesión Grupo Focal.....	79
<b>Tabla 12.</b>	Aspectos por mejorar identificados en la sesión del Grupo Focal.....	80
<b>Tabla 13.</b>	Evaluadores de los modelos.....	81
<b>Tabla 14.</b>	Lineamientos de calidad para la evaluación de los modelos.....	81

<b>Tabla 15.</b> Evaluación del servicio de detección de similitud de propuestas .....	82
<b>Tabla 16.</b> Evaluación del servicio de detección de manipulación de ofertas .....	83
<b>Tabla 17.</b> Evaluación del servicio de detección de organización de carteles.....	83
<b>Tabla 18.</b> Evaluación del servicio de alerta ciudadana. ....	84

## Índice de Figuras

<b>Figura 1.</b> Estudios primarios por año.....	15
<b>Figura 2.</b> Diagrama macro de componentes del modelo. ....	19
<b>Figura 3.</b> Descripción general del proceso de detección de similitud .....	24
<b>Figura 4.</b> P.N. del módulo de detección de similitudes .....	27
<b>Figura 5.</b> P.N. del servicio de detección de probable manipulación de ofertas ...	36
<b>Figura 6.</b> Pseudocódigo algoritmo LexRank adaptado de [56] .....	42
<b>Figura 7.</b> Beneficiarios finales.....	49
<b>Figura 8.</b> Proceso de recopilación de la fuente de información del servicio .....	53
<b>Figura 9.</b> P.N. del servicio de detección de probable organización de carteles....	54
<b>Figura 10.</b> Gráfica de interacciones de proponentes. Tomada de [65] .....	55
<b>Figura 11.</b> Diagrama de procesos de negocio servicio de denuncia ciudadana. ...	61
<b>Figura 12.</b> Arquitectura control de acceso basado en roles.....	64
<b>Figura 13.</b> Flujo de administración, autenticación y autorización de usuarios .....	64
<b>Figura 14.</b> Diagrama de secuencia que ejemplifica una petición en MVC.....	66
<b>Figura 15.</b> Diagrama general del servicio de similitud en propuestas técnicas. ...	70
<b>Figura 16.</b> Pantalla: Listado de convocatorias. ....	73
<b>Figura 17.</b> Pantalla: Listado de propuestas técnicas. ....	73
<b>Figura 18.</b> Pantalla: Resultado del análisis de similitud.....	74
<b>Figura 19.</b> Listado de propuestas económicas de una convocatoria. ....	75
<b>Figura 20.</b> Resultado del análisis de manipulación de ofertas (3 propuestas participes del delito). ....	76
<b>Figura 21.</b> Resultados del análisis de similitud escenario copiado textual. ....	85
<b>Figura 22.</b> Resultados del análisis de similitud escenario copiado inteligente.....	86
<b>Figura 23.</b> Resultados del análisis de similitud escenario sin copia intencional. ...	86
<b>Figura 24.</b> Gráfica de evaluación valores del umbral de detección de similitud....	87
<b>Figura 25.</b> Matriz de confusión y métricas sobre la experimentación del servicio de detección de similitud en propuestas técnicas. ....	87
<b>Figura 26.</b> Resultados del análisis de manipulación de ofertas escenario de solo un documento partícipe del delito.....	88
<b>Figura 27.</b> Resultados del análisis de manipulación de ofertas escenario de solo un documento partícipe del delito.....	89
<b>Figura 28.</b> Gráfica de afinamiento del umbral de detección de manipulación de ofertas. ....	89
<b>Figura 29.</b> Matriz de confusión y métricas sobre la experimentación del servicio de detección de probable manipulación de ofertas. ....	90



## Índice de Anexos

**Anexo 1:** Prototipo Software y código fuente de la herramienta desarrollada como prueba de concepto del modelo propuesto.

**Anexo 2:** Artículo “Una revisión sistemática de los métodos de identificación y detección de corrupción en contratación pública”.

**Anexo 3:** Artículo “Modelo para definir índices de riesgo de corrupción en convocatorias de contratación estatal en Colombia basada en Big Data y Procesamiento del Lenguaje Natural”.

# CAPÍTULO 1

---

## 1. Introducción

### 1.1 Planteamiento del problema

El origen etimológico de la palabra corrupción viene del latín corruptio, que hace referencia a los cambios abruptos con respecto a lo que se considera la naturaleza esencial de las cosas. Con esto, se puede entender que la corrupción conlleva un cambio en la esencia de algo, que deja de ser lo que era o se distancia de su intención original. En la actualidad, la corrupción se define como una práctica deshonesta llevada a cabo por individuos, que abusan del poder público o privado para su beneficio propio o de alguna organización (empresa comercial, partido político, entre otros).

En Colombia la percepción de la corrupción se ha elevado en los últimos años, según lo revela el Índice de Percepción de Corrupción (IPC) de la agencia para la Transparencia Internacional en el año 2021 [1]. Este informe califica en un rango de 0 a 100 la percepción sobre corrupción en el país, obteniendo un resultado promedio de 39 puntos, el cual está por debajo del promedio regional que es de 45 puntos, esto hace que el país se ubique en el puesto 18 con respecto a la región y el puesto 87 entre 180 países del mundo. En el 2021 según datos entregados por el contralor general de Colombia Carlos Felipe Córdoba, se estima que el flagelo de la corrupción le cuesta al país el equivalente al 4 % del Producto Interno Bruto (PIB) o, lo que es lo mismo que el 17% del presupuesto general de la nación (que para el año 2021 fue de \$303 billones). Lo que corresponde a 51,5 billones de pesos al año, casi un billón de pesos semanales. Así mismo menciona que este monto sería suficiente para atender el pago del servicio de la deuda de la Nación en ese año (54 billones). Similarmente con ese dinero se podría llevar a cabo cuatro líneas del metro en la ciudad de Bogotá (teniendo en cuenta que el contrato para la primera línea es de trece billones de pesos), o mejor aún, se cubriría el costo estimado de las llamadas obras de Cuarta Generación, unos 45 billones de pesos.

En un país como Colombia, un estado social de derecho, el Estado debe priorizar la satisfacción del interés general y el proceso de Contratación Estatal es una herramienta para el logro de este objetivo, proceso que cumple una serie de principios y reglas. La corrupción como un fenómeno propio de la contratación estatal es uno de los problemas sociales con mayor trascendencia en la sociedad actual. Cuando ocurre un caso de corrupción, el proceso de contratación deja de estar dirigido a la satisfacción del interés general y se enmarca en la satisfacción de intereses personales o de pequeños grupos de agentes del proceso contractual, lo que afecta el bienestar de la comunidad [2].

Según la OCDE la contratación pública representa aproximadamente el 15% del producto interno bruto (PIB) en los países pertenecientes a esta organización [3], [4]. Por su parte Transparencia Internacional determinó que aproximadamente el 70% del presupuesto Estatal de un país en desarrollo se ejecuta por medio de contratos, convirtiendo, el proceso de contratación estatal en un probable foco de corrupción. Según [2] la corrupción en la contratación estatal tiene como ingredientes el soborno y la recompensa ilegal (incentivos ilegales para manipular las decisiones de los servidores públicos en beneficio propio).

El aumento de los casos de corrupción ha hecho que diversas entidades estén aplicando variadas estrategias anticorrupción, desafortunadamente la mayoría de estas son técnicas manuales y de enfoque social, que buscan prevenir, más que identificar probables casos de corrupción. Si bien las técnicas manuales pueden ser útiles como estrategias anticorrupción, se hace necesario desarrollar estrategias automáticas (la identificación manual de la corrupción es intrínsecamente difícil [5]) y predictivas (basadas en estadística, inteligencia artificial y otras técnicas) para reducir y en un futuro eliminar este flagelo de la sociedad.

En la literatura (ver detalles más adelante en la **sección 2** relacionada con el Estado del Arte) se pueden encontrar diversas estrategias que se han utilizado para la identificación de presuntos delitos asociados con la corrupción, mediante la evaluación de un delito específico, como lo son investigaciones basadas en la teoría del diamante para prevenir e identificar corrupción en contratos públicos [6], [7]. Además de artículos que hacen uso de diversas técnicas estadísticas como Test econométricos [8], [9], pantallas [10], regresión [11] y teoría de conjuntos difusos [12], [13]. Más recientemente se presentan estrategias híbridas entre estadística e Inteligencia artificial para atacar este problema, destacándose la utilización de la teoría de conjuntos difusos con algoritmos de agrupamiento [14] o técnicas de evaluación difusas [15], así mismo la combinación del análisis de regresión con otros algoritmos predictivos [16] o con Big Data [17] y finalmente el uso de pantallas estadísticas con aprendizaje de máquina [18]. En los últimos años también se destaca el uso de algoritmos de agrupamiento como el K-means [19] combinados con reglas de asociación [20] en un enfoque de minería de datos. Y el uso de redes neuronales como los mapas autoorganizados (redes de Kohonen) [21], además de técnicas basadas en el análisis de grafos [22] con información de las direcciones IP de los proponentes en una licitación [23].

Es de destacar que las investigaciones mencionadas anteriormente buscan resolver el problema de identificación de presuntas situaciones de corrupción en procesos de contratación haciendo énfasis en un delito en específico, mayormente colusión, creación de carteles o manipulación de ofertas, esto mediante el uso, por lo general, de una o dos herramientas para ello. Sin embargo, los estudios también muestran que, dada la complejidad del problema, se hace necesario contar con un conjunto (batería) de herramientas especializadas para atacar los diferentes delitos que ocurren en el proceso de contratación, dado esto, se logra entender la importancia que tiene continuar con el desarrollo de modelos que permitan solventar de manera más efectiva (eficaz y eficiente) el problema de la identificación de presuntas

situaciones de corrupción en la contratación. Además, estos trabajos han sido desarrollados en otros países y no en Colombia.

Finalmente, en modelos de índole predictivo de clasificación o estimación (regresión) se hace necesario contar con una tasa de aciertos mínima (de acuerdo con la medida de calidad que sea la más apropiada al problema que se busca resolver, por ejemplo en problemas de clasificación pueden ser la exactitud, precisión, recuerdo, medida F, AUC, entre otras y en problemas de estimación pueden ser la suma de los errores cuadrados, el coeficiente de determinación, entre otros) para que el modelo sea usable y no genere más trabajo del deseado, por ejemplo, al obtener muchos falsos positivos, casos predichos como presunto delito y que en la realidad no lo son, o que por el contrario sean muy flexibles con situaciones que en la realidad son delitos y la herramienta los prediga como que no los son. Aunque la literatura no reporta un valor absoluto aceptable para todos los casos y no se encuentra dicha métrica en las investigaciones previas de corrupción se buscó en esta investigación obtener una tasa igual o superior a 80% en las medidas relacionadas con problemas de clasificación (exactitud, precisión, recuerdo y medida F).

Teniendo en cuenta lo anterior, en el presente trabajo se partió de la siguiente pregunta de investigación: ¿Qué servicios especializados debe incluir un modelo conceptual basado en Big Data y Procesamiento del Lenguaje Natural que permita identificar presuntas situaciones de corrupción en los datos y documentos asociados a los procesos de contratación pública en Colombia? Y para que este conjunto de servicios especializados pueda ser luego implementado, se buscó establecer las entradas, el proceso que deberían realizar, los soportes tecnológicos que se podrían usar y las salidas que cada servicio debería entregar.

Con la anterior pregunta, la hipótesis principal de la investigación fue:

H1: Incluir un conjunto de servicios especializados dentro de un modelo conceptual anticorrupción basado en Big Data y Procesamiento del Lenguaje Natural permite obtener resultados con calidad superior al 80% en la identificación de presuntas situaciones de corrupción en los datos y documentos asociados a los procesos de contratación pública.

Por otro lado, la hipótesis alternativa planteada para esta investigación fue:

H0: incluir un conjunto de servicios especializados dentro de un modelo conceptual anticorrupción basado en Big Data y Procesamiento del Lenguaje Natural NO permite obtener resultados con calidad superior al 80% en la identificación de presuntas situaciones de corrupción en los datos y documentos asociados a los procesos de contratación pública.

Teniendo en cuenta que esta pregunta de investigación tiene un amplio alcance y que un trabajo de grado en modalidad de investigación en el programa de maestría en computación en la Universidad del Cauca se debe realizar máximo en doce (12) meses, se hizo necesario definir un enfoque de solución que permitiera acotar el

alcance de la tesis. Para lograr eso, se realizó primero una revisión sistemática de la literatura para conocer los modelos de solución que se han desarrollado hasta el momento, buscando resolver el problema de la identificación de corrupción en el proceso de contratación, luego, se procedió a realizar un modelo conceptual que integra diferentes servicios para la identificación de presuntas situaciones de corrupción, modelo que fue evaluado por parte de expertos en el área a través de grupos focales, y finalmente, la implementación de dos (2) servicios especializados a modo de prueba de concepto que fueron evaluados con base en la calidad (exactitud, precisión, recuerdo, medida F1 u otra) de la identificación de presuntas situaciones de corrupción para los cuales fueron diseñados con base en datos sintéticos. Con base en estos resultados se espera que se puedan iniciar futuros procesos de adaptación y uso de estos servicios y nuevas implementaciones y mejoras que se puedan integrar en el modelo para que este se desarrolle en el tiempo con aportes de la comunidad académica, científica e industrial.

## **1.2 Aportes del proyecto**

Desde el punto de vista de investigación, las contribuciones del presente proyecto están enfocadas en la generación de nuevo conocimiento relevante a la definición de índices de riesgo de presuntas situaciones de corrupción en el proceso de contratación basado en técnicas de inteligencia artificial y de procesamiento de lenguaje natural. Propuesta que es nueva y diferente teniendo en cuenta que hoy en día no hay investigaciones publicadas (en las bases de datos de Scopus, ACM, IEEE, ScienceDirect y SpringerLink) donde se evidencie el uso de un modelo marco o general, que haga uso de diferentes técnicas para identificar diversos tipos de presuntas situaciones de corrupción.

Desde el punto de vista de innovación y desarrollo, hay dos aportes, el primero se centra en la entrega de unos servicios ya implementados con base en el modelo propuesto, estos pueden pasar a producción en diversas agencias gubernamentales que buscan prevenir la corrupción o agencias no gubernamentales que advierten (análisis post mortem) la corrupción en el país, para lograr esto, es claro que se requerirán adaptaciones según los datos y el proceso que estas instituciones manejen, pero partirían ya de un producto que cumple con objetivos bien definidos y acordes con la tarea. El segundo se centra en que con el modelo propuesto y los servicios que falta por desarrollar quedan oportunidades para dar continuidad al trabajo desarrollado, a través de futuros trabajos de grado a nivel de pregrado o maestría o trabajos empresariales que terminen la implementación del modelo y lo implementen en agencias anticorrupción.

Vale la pena mencionar que la herramienta resultante servirá como guía, a manera de alerta temprana para identificar presuntas situaciones de corrupción, y es el tomador de decisiones del proceso de contratación quien al final debe juzgar o no si la situación amerita iniciar un proceso disciplinario, civil, penal o combinaciones de estos tipos. El modelo y los servicios desarrollados no pretenden juzgar una propuesta, sino soportar la toma de decisiones.

### 1.3 Objetivos

A continuación, se presentan los objetivos tal y como fueron aprobados en el documento de anteproyecto por parte del Consejo de Facultad de la Facultad de Ingeniería Electrónica y Telecomunicaciones de la Universidad del Cauca.

#### 1.3.1 Objetivo general

Proponer un Modelo basado en Big Data y procesamiento de lenguaje natural para definir índices de riesgo de presuntas situaciones de corrupción en los datos y documentación asociada a las convocatorias de contratación estatal en Colombia.

#### 1.3.2 Objetivos específicos

1. Establecer una revisión sistemática de los modelos, estrategias o técnicas utilizadas para la identificación de corrupción en el proceso de contratación haciendo uso de las guías propuestas por Kitchenham y otros en [24], buscando identificar los enfoques más recientes y exitosos de solución y establecer un mapa de los posibles usos/aplicaciones de las técnicas que a la fecha no se han usado en este contexto.
2. Definir un modelo que integre técnicas de Big Data y procesamiento de lenguaje natural para definir índices de riesgo de presuntas situaciones de corrupción en los datos y documentos relacionados con contratación estatal en Colombia haciendo un análisis de los posibles delitos que se pueden presentar, guiado por el patrón de investigación iterativa propuesto por Pratt [25] y representado en diagramas estructurales (componentes), de comportamiento (actividades) y de interacción (secuencia y procesos) definidos en UML.
3. Desarrollar una herramienta software, como prueba de concepto, implementada en el lenguaje de programación Python usando como metodología de desarrollo SCRUM, que soporte un máximo de dos servicios especializados del modelo propuesto y que permita evaluar la calidad de la predicción obtenida en dichos servicios, mediante métricas clásicas del área (exactitud, precisión, recuerdo, medida F1 u otras) en un conjunto de datos sintéticos de evaluación.

### 1.4 Resultados obtenidos

A continuación, se resumen los principales resultados de la presente investigación:

- **Monografía de la investigación.** Corresponde al presente documento y resume el problema, los objetivos planteados para resolver dicho problema, los productos obtenidos (modelo y servicios desarrollados) así como la evaluación realizada sobre el trabajo desarrollado. Finalmente, presenta las conclusiones y unas ideas de trabajo futuro que sería de interés desarrollar en el área de investigación.
- **Modelo propuesto e implementación de dos servicios,** junto con su documentación y código fuente.

- **Dos artículos:** resultado de la investigación realizada. El primero corresponde al resultado de una revisión de la literatura que ha sido publicado en una revista categoría B del PUBLINDEX de Minciencias con la siguiente citación: Julio-César Luna-Ortega, Carlos-Alberto Cobos-Lozada, Martha-Eliana Mendoza-Becerra. A Systematic Review of the Methods for Identification and Detection of Corruption in Public Procurement. RISTI - Revista Ibérica de Sistemas e Tecnologías de Informação, E38, 43-57. Diciembre 2020, ISSN: 1646-9895. En el segundo artículo se plasman los resultados completos de la investigación, el cual se anexa a esta monografía y actualmente se encuentra en proceso de evaluación en un evento internacional con memorias en revista mínimo categoría B según el PUBLINDEX de COLCIENCIAS.

## 1.5 Estructura de la monografía

A continuación, se describe de manera general el contenido y organización de la presente monografía:

**CAPITULO 1: INTRODUCCIÓN:** Hace referencia al presente capítulo que introduce el tema de investigación, presenta la pregunta de investigación que originó el trabajo, los aportes al problema, también los objetivos (general y específicos) definidos para el proyecto, un breve resumen de los resultados obtenidos y finalmente la organización de la monografía.

**CAPITULO 2: ESTADO DEL ARTE:** Este capítulo presenta el estado del arte de trabajos relacionadas con identificación de corrupción a través de una revisión sistemática de la literatura. De esta revisión se muestran los pasos seguidos y el resultado obtenido, junto con las conclusiones de esta.

**CAPITULO 3: MODELO PROPUESTO:** En este capítulo se presenta el modelo propuesto en esta investigación, así como los diferentes servicios que componen al mismo, junto con la revisión de la literatura de cada uno de estos servicios.

**CAPITULO 4: SERVICIOS DESARROLLADOS:** En este capítulo se presenta el diseño general y las interfaces más importantes junto con su descripción, de los dos servicios especializados desarrollados en el marco del modelo propuesto.

**CAPITULO 5: EVALUACIÓN DEL MODELO Y LOS SERVICIOS:** Aquí se presenta la evaluación del modelo a través de grupos focales y de los dos servicios desarrollados a través de datos sintéticos y medidas clásicas de evaluación.

**CAPITULO 6: CONCLUSIONES Y TRABAJO FUTURO:** En este capítulo se presentan las conclusiones obtenidas al finalizar el trabajo de grado e ideas que el grupo de investigación y terceros pueden realizar como trabajo futuro en la temática de interés de esta investigación.

**CAPITULO 7: BIBLIOGRAFIA:** Este último capítulo contiene las referencias bibliográficas de los artículos y libros consultados para la realización del proyecto.

## CAPÍTULO 2

---

### 2. Estado del Arte

Para definir el estado del arte se realizó una Revisión Sistemática de la Literatura (RSL) del tema en cuestión siguiendo la guía propuesta en [26]. Esta revisión se postuló y después de su evaluación por pares internacionales fue aceptada como ponencia en el evento denominado JIISIC 2020 – XV Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería del Conocimiento, que se realizó virtualmente del 19 al 21 de agosto. Finalmente, los organizadores del evento definieron que la publicación se realizará en una revista categoría B (SJR Q3) del PUBLINDEX de Minciencias, con la siguiente referencia: Julio-César Luna-Ortega, Carlos-Alberto Cobos-Lozada, Martha-Eliana Mendoza-Becerra. A Systematic Review of the Methods for Identification and Detection of Corruption in Public Procurement. RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação, E38, 43-57. Diciembre 2020, ISSN: 1646-9895. A continuación, se presenta el resultado de este trabajo. Luego de realizada esta RSL se han realizado nuevos hallazgos relacionados con el modelo propuesto y cada uno de sus servicios, estos hallazgos se indican al inicio de la explicación de cada servicio en el siguiente capítulo.

#### 2.1 Planeación de la Revisión Sistemática de la Literatura (RSL)

Los documentos relacionados con la temática de investigación se buscaron en Scopus, ACM Digital Library, IEEEExplore y ScienceDirect. Posterior a esto, se aplicaron los criterios de inclusión/exclusión para la selección de los estudios primarios y finalmente se realizó una evaluación de la calidad de los estudios incluidos.

##### 2.1.1 Preguntas de investigación

Las preguntas de investigación de esta revisión buscan recopilar y categorizar la información existente acerca de la detección de corrupción en contratación pública, con el fin de realizar una revisión general del tema, identificando las carencias existentes y con esto, en un próximo paso plantear una alternativa de solución a la problemática expuesta (ver **Tabla 1**).

**Tabla 1.** Preguntas de investigación

Preguntas de Investigación	Motivación
Q1. ¿Qué estudios existen sobre la detección o identificación de corrupción en contratación?	Determinar el número de publicaciones actuales y la tendencia a lo largo de los últimos años en relación con el campo de la identificación de corrupción en contratación.



<b>Q2.</b> ¿Qué tipos de soluciones se han implementado para la identificación de corrupción?	Determinar cuáles son las técnicas, métodos, modelos o similares que se han utilizado para la detección de corrupción en contratación y como son los informes que estas realizan.
<b>Q3.</b> ¿Qué tipos de corrupción en contratación se han trabajado?	Determinar qué tipos de corrupción o fraude son normalmente a los que se presta más atención en el proceso de contratación.
<b>Q4.</b> ¿En qué dominios de aplicación se ha tratado el problema de la identificación de la corrupción?	Determinar cuáles son los objetos (contratos, negocios, etc.) en donde se ha trabajado sobre el problema de la identificación de corrupción.
<b>Q5.</b> ¿Qué indicadores se utilizan en las mediciones de corrupción o fraude?	Determinar cuáles son los indicadores que se suelen utilizar a la hora de evaluar un contrato y cuál es el dominio de valores que estos pueden tomar para alertar un posible caso de corrupción.
<b>Q6.</b> ¿Qué problemas y desafíos identifican los autores en relación con la corrupción?	Determinar cuáles son los problemas y retos identificados por los autores que trabajan en el proceso de identificación de la corrupción en la contratación.

## 2.1.2 Desarrollo del protocolo de revisión

A continuación, se resume el protocolo seguido para llevar a cabo la RSL.

**Cadena de búsqueda de estudios primarios:** Los términos de la cadena de búsqueda fueron construidos a través de un proceso iterativo de mejora con el análisis de los resultados reportados por cada versión previa de esta. Se hizo uso de la estrategia planteada en [27], la cual consiste en derivar los términos de las preguntas de investigación, identificando y agrupándolos en población, intervención, comparación y resultado (ver **Tabla 2**). Los operadores booleanos OR incorporan palabras alternativas y sinónimos en cada una de las partes, y los operadores AND son utilizados para vincular los términos principales entre sí.

**Tabla 2.** Cadena de búsqueda usada para iniciar la revisión

Concepto	Estructura	Términos Alternativos y Sinónimos	Descripción
Técnicas	Resultado	("model" OR "algorithm" OR "framework" OR "technology" OR "method" OR "technique") AND	Se delimita lo que se está buscando: algoritmos, técnicas, modelos y similares.
Corrupción	Población	("*corruption" OR "collusion" OR "fraud") AND	Se define el problema a nivel general: corrupción y similares.
Tipos de delitos	Comparación	("bribery" OR "buys influence" OR "conflicts of interest" OR "cronyism" OR "embezzlement" OR "extort*" OR "favoritism" OR "gift-giving" OR "graft" OR "influence peddling" OR "kickback" OR "nepotism" OR "parochialism" OR "patronage" OR "political interference" OR "rigging" OR "splitting" OR "theft" ) AND	Se delimitan los tipos específicos de delitos de corrupción.
Objeto	Población	("government*" OR "agreement" OR "contract" OR "commitment" OR "arrangement" OR "settlement" OR "understanding" OR "bond" OR "deal" OR "bargain") AND	Se menciona el objeto que tiene el problema: contratación gubernamental y palabras relacionadas con contrato.
Acciones	Intervención	("detect" OR "avoid" OR "prevent" OR "warn" OR "caution" OR "identify" OR "understanding" OR "avert" OR "stop" OR "alert" OR "perceive" OR "distinguish" OR "notify" OR "inform" OR "recognize" OR "establish")	Se relaciona con el tipo de soluciones que buscan identificar, evitar, prevenir (o similares) la corrupción.

\*Los asteriscos significan que cualquier carácter o caracteres pueden ser incluidos en la palabra, de esta manera se consiguen permutaciones en los términos de búsqueda en los que se utiliza.

**Proceso de búsqueda:** Las búsquedas se realizaron sobre Scopus, ACM Digital Library, IEEEExplore y ScienceDirect. Siguiendo el proceso propuesto en [27], la búsqueda se organizó en fases; en la primera se identificaron los estudios candidatos derivados de las palabras claves de la cadena de búsqueda, una vez esto que se procedió a aplicar los criterios de selección sobre el abstract de cada uno de ellos, posteriormente los mismos criterios fueron aplicados a todo el documento de los estudios resultantes de la fase previa, finalmente se aplicó la técnica “bola de nieve”, la cual consiste en revisar las referencias de cada estudio primario buscando identificar más estudios candidatos, así mismo se hizo una revisión de la literatura gris. La **Tabla 3** resume el proceso de búsqueda realizado.

**Tabla 3.** Proceso de búsqueda

Actividad	Resultados
Aplicar cadena de búsqueda	245
Eliminar duplicados	229
Aplicar criterios de selección (inclusión y exclusión)	40
Leer documento completo	8
Revisar artículos excluidos y Snowballing (bola de nieve)	16
Revisar literatura gris	19
Aplicar criterios de evaluación de calidad	19

**Criterios de inclusión y exclusión:** Los criterios para incluir un estudio primario fueron los siguientes: 1) Estudios escritos en el idioma inglés; 2) Estudios completos publicados en revistas, conferencias, congresos o talleres de prestigio con revisión por pares; y 3) El estudio propone un modelo, técnica, algoritmo, framework o similar. Los criterios para excluir un estudio primario fueron los siguientes: 1) El artículo no presenta un diseño de investigación bien definido; 2) El artículo no presenta una pregunta de investigación bien definida; 3) El artículo corresponde a un estudio de debate, o solo está disponible en forma de resumen o presentación. 4) El artículo es un estudio secundario, es decir, corresponde a un reporte de resultados de otras investigaciones; 5) La contribución principal del artículo no se relaciona con alguna modelo/técnica/método/algoritmo de identificación de corrupción; 6) El artículo no propone un modelo/técnica/método/algoritmo, o que lo hagan, pero la propuesta no pueda ser implementada con ayuda del computador; y 7) El artículo presenta un modelo de índole psicosocial. Por último, vale la pena mencionar que, debido a la poca existencia de investigaciones que abordan el problema presentado en esta investigación, no se tuvo en cuenta el año de publicación del artículo como un criterio para incluir o no el trabajo en la presente revisión sistemática.

**Evaluación de calidad de los estudios primarios:** Adicional a la selección (criterios de inclusión y exclusión) se realizó una evaluación de la calidad de los estudios primarios [27] para ponderar la importancia de cada estudio seleccionado. Esta evaluación permitió comprender mejor las comparaciones entre los estudios y guiar la interpretación de los hallazgos. Con este objetivo se desarrolló una prueba con un sistema de puntuación de tres valores (-1, 0, 1) en 6 preguntas (ver **Tabla 4**). La puntuación total máxima corresponde a la suma de las calificaciones parciales

y el puntaje total está en el dominio [-6, 6]. El puntaje de calidad de cada estudio se presenta en la **Tabla 5**, donde 6 es el máximo valor posible.

**Tabla 4.** Prueba de evaluación de la calidad

No	Criterios	Puntuación		
		-1	0	1
C1	El estudio presenta una descripción detallada sobre los tipos de corrupción o fraude que se presentan en un objeto de población (ejemplo contratos)	No	Parcialmente	Si
C2	El estudio contiene un modelo/técnica/método/algorithmo/similar aplicada.	No	Parcialmente	Si
C3	El estudio tiene como finalidad aplicar acciones (identificar, detectar, prevenir, etc.) sobre un problema de corrupción	No	Parcialmente	Si
C4	El estudio expone de manera clara y detallada los resultados obtenidos tras aplicar un modelo/técnica/método/algorithmo/similar sobre un problema de corrupción en un objeto de población	No	Parcialmente	Si
C5	A qué cuartil pertenece la revista, conferencia o congreso en donde fue publicado el estudio	Q4 O NO	Q3	Q2 O Q1
C6	El estudio ha obtenido dos o más citas por año desde su publicación (según índice de citas de Google, Promedio anual)	menos de 1 cita por año	1 cita por año	2 o más citas por año

**Tabla 5.** Estudios primarios ordenado por año y puntaje

Referencia	(Porter & Zona, 1993) [8]	(Bajari, 2001) [28]	(Bajari & Ye, 2003) [12]	(Parente, Schulte, Jost, Sullivan, & Klindworth, 2012) [16]	(Choitibhongs & Arditi, 2012) [11]	(Ralha & Silva, 2012) [20]	(Morozov & Podkolzina, 2013) [9]	(Shan, Chan, Le, Xia, & Hu, 2015) [13]	(Fazekas, Tóth, & King, 2016) [17]	(Peltier-Rivest, 2017) [6]	(Belokrylov & Runova, 2017) [14]	(Islam & Yousuf, 2018) [19]	(Peltier-Rivest, 2018) [7]	(López-Iturriaga & Sanz, 2018) [21]	(Imhof, Karagök, & Rutz, 2018) [10]	(Wachs & Kertész, 2019) [22]	(Huber & Imhof, 2019) [18]	(Owusu, Chan, & Ameyaw, 2019) [15]	(Dorofeeva & Tokmakov, 2020) [23]
Puntaje	6	6	6	3	6	6	3	3	5	1	3	1	3	5	6	5	6	6	1

## 2.2 Resultados de la RSL

Haciendo uso del protocolo mencionado se encontraron 19 estudios que cumplieron con los diferentes criterios de calidad establecidos. Estos estudios se analizaron y organizaron en 5 grupos (A, B, C, D y E) conforme se presentan en la **Tabla 6**. En esta tabla, además, se muestra para cada estudio las respuestas a cada pregunta de investigación previamente formulada. Para determinar el grupo al que pertenecía cada estudio se hizo un análisis de las respuestas obtenidas por cada uno de ellos en las diferentes preguntas de investigación, siendo la pregunta **Q2** la de mayor peso en esta decisión.

**Tabla 6.** Agrupamiento de los estudios

<b>Id</b>	<b>Ref.</b>	<b>Preguntas</b>
A	[8]	<p><b>Q1:</b> Propuesta.  <b>Q2:</b> Caso de estudio de un cartel con información de cuentas debidamente detalladas respecto a sus actividades y prácticas de licitación. Técnicas estadísticas. Procedimientos de prueba econométricos.  <b>Q3:</b> Colusión. Manipulación de ofertas (práctica ilegal en la que los competidores se confabulan para elegir al ganador de una licitación).  <b>Q4:</b> Contrato.  <b>Q5:</b> No utiliza.  <b>Q6:</b> Falta de información de contratación pública.</p>
	[12]	<p><b>Q1:</b> Propuesta.  <b>Q2:</b> Técnicas estadísticas. Enfoque para identificar y probar la manipulación de ofertas en subastas de adquisición. Los autores destacan la importancia de la implementación de herramientas computacionales que puedan ser utilizadas para detectar la manipulación de ofertas.  <b>Q3:</b> Manipulación de ofertas.  <b>Q4:</b> Contrato.  <b>Q5:</b> Índice de probabilidad.  <b>Q6:</b> Falta de información de contratación pública. Empresas corruptas pueden aplicar acciones que ayuden a evitar la identificación de la corrupción.</p>
	[11]	<p><b>Q1:</b> Propuesta.  <b>Q2:</b> Técnicas estadísticas. Método paso a paso para detectar las ofertas colusorias mediante el análisis de los datos históricos de ofertas proporcionados por el propietario de una construcción. El método se basa en un modelo de regresión e implica identificar a los probables postores del cártel utilizando la prueba residual y la prueba de estabilidad de la estructura de costos.  <b>Q3:</b> Colusión. Manipulación de ofertas.  <b>Q4:</b> Contrato.  <b>Q5:</b> Niveles de correlación.  <b>Q6:</b> La corrupción en los proyectos es un problema complejo para el que no hay una fácil o única solución. La corrupción puede tomar muchas formas. El modelo no puede detectar las ofertas colusorias en el momento en que sucede; se usa para detectar probable colusión en el pasado y solo si hay acceso a información histórica detallada sobre un conjunto de variables independientes.</p>
	[9]	<p><b>Q1:</b> Propuesta.  <b>Q2:</b> Técnicas estadísticas. Método econométrico de detección de manipulación de ofertas.  <b>Q3:</b> Colusión. Manipulación de ofertas.  <b>Q4:</b> Contrato.  <b>Q5:</b> Niveles de correlación.  <b>Q6:</b> La corrupción en los proyectos es un problema complejo para el que no hay una fácil o única solución.</p>
	[13]	<p><b>Q1:</b> Encuesta. Identificación de elementos importantes de medición de corrupción.  <b>Q2:</b> Técnicas estadísticas. Modelo sistemático para medir la corrupción potencial en un proyecto de construcción pública.  <b>Q3:</b> No enfatiza un delito.  <b>Q4:</b> Contrato.  <b>Q5:</b> No utiliza.  <b>Q6:</b> Falta de información de contratación pública.</p>
	[10]	<p><b>Q1:</b> Propuesta.  <b>Q2:</b> Técnicas estadísticas. Método para detectar la manipulación de ofertas mediante la aplicación de pantallas (índices estadísticos específicos). El método es particularmente adecuado para detección de la colusión parcial (colusión que no involucra a todas las empresas y/o todos los contratos en un conjunto de datos específico). Se logra aislar a un grupo de empresas "sospechosas" de ser un cartel local de manipulación de ofertas.  <b>Q3:</b> Colusión. Manipulación de ofertas.  <b>Q4:</b> Contrato.  <b>Q5:</b> Índice de probabilidad.  <b>Q6:</b> Falta de información de contratación pública. La corrupción en los proyectos es un problema complejo para el que no hay una fácil o única solución. La corrupción puede tomar muchas formas.</p>
	B	[16]

	<p><b>Q1:</b> Propuesta.  <b>Q2:</b> Técnicas estadísticas. Inteligencia artificial. El método hace uso de Big Data para crear un indicador de corrupción a nivel de contrato.  <b>Q3:</b> No enfatiza un delito.  <b>Q4:</b> Contrato.  <b>Q5:</b> Banderas rojas. Propone indicador de corrupción a través de Big Data.  <b>Q6:</b> La corrupción en los proyectos es un problema complejo para el que no hay una fácil o única solución.</p>
[17]	<p><b>Q1:</b> Propuesta.  <b>Q2:</b> Técnicas estadísticas. Algoritmos computacionales. Propone un modelo basado en lógica difusa.  <b>Q3:</b> Manipulación de ofertas. Sobornos.  <b>Q4:</b> Contrato.  <b>Q5:</b> No utiliza.  <b>Q6:</b> La corrupción en los proyectos es un problema complejo para el que no hay una fácil o única solución. Falta de compromiso de las entidades.</p>
[18]	<p><b>Q1:</b> Propuesta.  <b>Q2:</b> Técnicas estadísticas. Inteligencia artificial Combinan técnicas de aprendizaje automático con pantallas estadísticas.  <b>Q3:</b> Manipulación de ofertas.  <b>Q4:</b> Contrato.  <b>Q5:</b> Binarios. Clasificación binaria que informa si una oferta es marcada como posible colusoria, clasificando como 1 si la probabilidad de colusión de la oferta es superior a 0,5 y 0 de lo contrario.  <b>Q6:</b> Debe realizarse de forma discreta, si algún cartel se entera del proceso, podría destruir evidencia. Falta de compromiso de las entidades. Información de costos es difícil de obtener.</p>
[15]	<p><b>Q1:</b> Propuesta. Muestra una revisión de la literatura con 21 actividades que se deben realizar en las cuatro etapas del proceso de adquisición.  <b>Q2:</b> Técnicas estadísticas. Algoritmos computacionales. Modelo que usa un enfoque de computación flexible (técnica de evaluación sintética difusa).  <b>Q3:</b> No enfatiza un delito.  <b>Q4:</b> Contrato.  <b>Q5:</b> Índice de probabilidad.  <b>Q6:</b> La corrupción en los proyectos es un problema complejo para el que no hay una fácil o única solución.</p>
[20]	<p><b>Q1:</b> Propuesta. Se enfoca en extraer información de las bases de datos del proceso de adquisiciones federales brasileñas para identificar carteles.  <b>Q2:</b> Inteligencia artificial. Minería de datos: Agrupamiento y reglas de asociación.  <b>Q3:</b> Carteles.  <b>Q4:</b> Contrato.  <b>Q5:</b> Binario.  <b>Q6:</b> Falta de compromiso de las entidades.</p>
C [19]	<p><b>Q1:</b> Propuesta.  <b>Q2:</b> Inteligencia artificial. Se ha propuesto un modelo que usa el algoritmo K-means.  <b>Q3:</b> No enfatiza en un delito.  <b>Q4:</b> Corrupción pública.  <b>Q5:</b> Índice de probabilidad.  <b>Q6:</b> Falta de compromiso de las entidades.</p>
[21]	<p><b>Q1:</b> Propuesta.  <b>Q2:</b> Inteligencia artificial Propone sistema de alertas tempranas basado en red neuronal (mapas autoorganizados, SOM).  <b>Q3:</b> No enfatiza un delito.  <b>Q4:</b> Corrupción pública.  <b>Q5:</b> Binarios.  <b>Q6:</b> Falta de compromiso de las entidades.</p>
[28]	<p><b>Q1:</b> Propuesta.  <b>Q2:</b> Algoritmos computacionales. Los algoritmos computacionales desarrollados en este documento pueden usarse para la evaluación de los daños causados por la manipulación de ofertas.  <b>Q3:</b> Colusión. Manipulación de ofertas.  <b>Q4:</b> Contrato.  <b>Q5:</b> Diferencia entre ofertas corruptas y competitivas.  <b>Q6:</b> La corrupción en los proyectos es un problema complejo para el que no hay una fácil o única solución. Falta de información de contratación.</p>
D [22]	<p><b>Q1:</b> Propuesta.  <b>Q2:</b> Ciencia de redes. Propone un marco basado en análisis de redes (grafos) para identificar grupos de empresas que interactúan intensamente en un mercado.  <b>Q3:</b> Colusión. Manipulación de ofertas.</p>

	<p><b>Q4:</b> Contrato.  <b>Q5:</b> Índices de probabilidad.  <b>Q6:</b> Falta de información de contratación pública. La corrupción puede tomar muchas formas.</p>
[23]	<p><b>Q1:</b> Propuesta.  <b>Q2:</b> Inteligencia artificial. Análisis de las direcciones IP de los miembros de una licitación pública.  <b>Q3:</b> Manipulación de ofertas.  <b>Q4:</b> Contrato.  <b>Q5:</b> No utiliza.  <b>Q6:</b> Falta de compromiso de las entidades. Falta de claridad en la definición del objeto a abordar.</p>
[6]	<p><b>Q1:</b> Estudio de caso.  <b>Q2:</b> No relacionado con tecnología o estadística. Propone un marco de gobernanza, haciendo uso de la teoría del diamante del fraude.  <b>Q3:</b> Manipulación de ofertas. Sobornos. Conflictos de interés. Extorsión.  <b>Q4:</b> Contrato.  <b>Q5:</b> Banderas rojas.  <b>Q6:</b> Falta de información de contratación pública. Falta de compromiso de las entidades.</p>
E	<p><b>Q1:</b> Propuesta. Muestra aspectos interesantes para tener en cuenta en el proceso de detección de la corrupción.  <b>Q2:</b> No relacionado con tecnología o estadística. Introduce un modelo derivado de las teorías criminológicas (como el triángulo de fraude).  <b>Q3:</b> Manipulación de ofertas. Sobornos.  <b>Q4:</b> Contrato.  <b>Q5:</b> Banderas rojas.  <b>Q6:</b> Debe realizarse de forma discreta, si algún cartel se entera del proceso, podría destruir evidencia. Falta de compromiso de las entidades.</p>

### 2.2.1 Análisis de los grupos resultantes

A continuación, se muestran un análisis detallado de la conformación de cada uno de los grupos con relación a las preguntas de investigación planteadas.

**Grupo A:** Los estudios pertenecientes a este grupo tienen un enfoque de solución relacionada con técnicas estadísticas como Test econométricos ([8], [9]), pantallas ([10]), regresión ([11]) y teoría de conjuntos difusos ([12], [13]). Los tipos de delitos de corrupción abordados por este grupo de estudios son en su mayoría la colusión y la manipulación de ofertas, así mismo las métricas de corrupción utilizadas tienden a ser análisis estadísticos como es el caso de un índice de probabilidad. Es de destacar que los estudios pertenecientes a este grupo se presentan desde 1993 hasta 2018.

**Grupo B:** La principal característica de los estudios pertenecientes a este grupo es que el modelo desarrollado involucra técnicas estadísticas con métodos de Inteligencia Artificial, destacándose la utilización de la teoría de conjuntos difusos con algoritmos de agrupamiento ([14]) buscando medir el riesgo de corrupción en el proceso de contratación o técnicas de evaluación difusas con el objetivo de evaluar los niveles de vulnerabilidad con respecto a las etapas de corrupción ([15]), así mismo la combinación del análisis de regresión con algoritmos predictivos para la detección analítica de fraude ([16]) o con algoritmos de Big Data ([17]) para el desarrollo de un indicador compuesto de corrupción y finalmente la combinación de pantallas estadísticas con aprendizaje de máquina ([18]) con el objetivo de predecir la colusión a través de carteles de manipulación de ofertas. La mayoría de estos estudios no enfatizan su solución a un delito en particular, y los que los hacen se enfocan en el de manipulación de ofertas. Por otro lado, similar al grupo anterior, la

tendencia de las métricas de corrupción utilizadas se encuentra en un marco estadístico. Estos trabajos son relativamente nuevos, teniendo su primera publicación en 2012, así mismo, se puede apreciar que la tendencia de este tipo de artículos ha venido incrementando en los últimos años.

**Grupo C:** Los integrantes de este grupo son estudios en donde se aplicó una solución tecnológica basada en Minería de Datos, haciendo uso de algoritmos de agrupamiento y reglas asociación. Como en [20], donde se obtiene un modelo que permitió detectar las ofertas colusorias mediante el análisis de los datos históricos de ofertas. Por su parte, en [19] hacen uso del algoritmo K-means (agrupamiento) el cual permitió crear diferentes grupos de acuerdo con el nivel de corrupción de un grupo de entidades. Finalmente, en [21] se usó un enfoque de red neuronal, mapas autoorganizados (agrupamiento), para predecir casos de corrupción en diferentes espacios temporales, en una región específica de un país. Los estudios pertenecientes a este grupo por lo general no se enfocan en un delito en particular, solo uno de ellos lo hace, enfocándose en la organización de carteles. Por otro lado, en este grupo prima un indicador binario como métrica de corrupción. Finalmente, y, como era de esperarse, similar al anterior grupo este tipo de soluciones es relativamente nueva, teniendo su primera publicación en 2012.

**Grupo D:** Los integrantes de este grupo, si bien hacen uso de técnicas computacionales, difieren del anterior en que dichas técnicas no están enmarcadas dentro de la Inteligencia Artificial. Destacándose soluciones que hacen uso de algoritmos computacionales ([28]) para encontrar diferencias entre ofertas competitivas y colusorias, o la utilización de análisis de redes (grafos) ([22]) para la identificación de grupos de empresas que interactúan intensamente en un mercado y detectar su potencial colusorio. Y finalmente el análisis de las direcciones IP de los miembros de una licitación pública para la identificación de posible presencia de un cartel en un proceso de licitación ([23]). Este grupo se centra en la detección de colusión y manipulación de ofertas, teniendo como principal métrica los índices de probabilidades. Finalmente, y similar a los anteriores grupos que hacen uso de técnicas de tecnologías de la información, la tendencia de este tipo de investigaciones es relativamente nueva.

**Grupo E:** El último grupo de investigaciones está conformado por los estudios cuyo modelo de solución no está relacionado con tecnología o estadística, en estos estudios se despliegan una serie de estrategias para prevenir o detectar la corrupción, haciendo uso de la teoría del diamante del fraude ([6], [7]). El delito que mayormente se aborda en este grupo es el de manipulación de ofertas, teniendo como métrica el uso de banderas rojas (indicador binario).

Por último, en la **Figura 1** se muestran los estudios primarios realizados por año, discriminados por la técnica utilizada, se logra apreciar que en los últimos años hay un crecimiento de los modelos que utilizan técnicas computacionales.

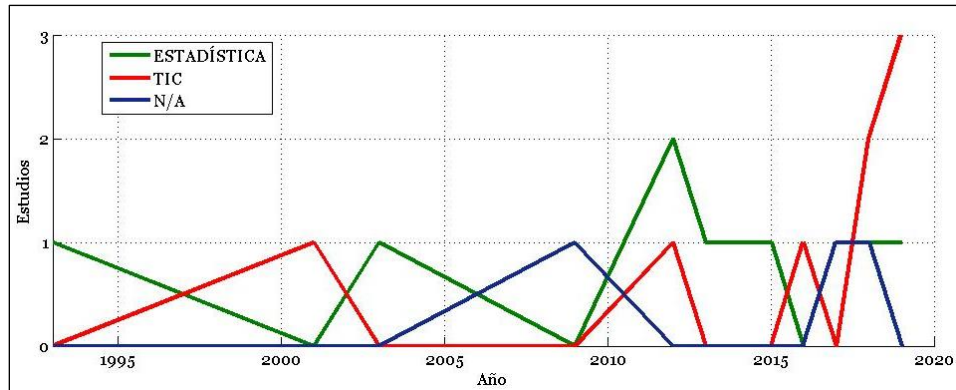


Figura 1. Estudios primarios por año

### 2.3 Conclusiones de la RSL

Los resultados obtenidos en esta revisión sistemática demuestran que el uso de técnicas estadísticas para solventar el problema de la identificación temprana de la corrupción es una idea que en los últimos años ha tenido una buena acogida, así como la combinación de estas técnicas con tecnologías de la información. Lo anterior sumado al incremento en los casos de corrupción global reportados, llevan a la necesidad de seguir en el desarrollo de modelos que permitan solventar de manera más eficaz y eficiente el problema de la detección de la corrupción en procesos de contratación. Los estudios muestran que, dada la complejidad del problema, se hace necesario contar con un conjunto (batería) de herramientas especializadas para atacar los diferentes delitos que ocurren en contratación.



Esta página ha sido intencionalmente dejada en blanco.

## CAPÍTULO 3

---

### 3. Modelo propuesto

Un modelo conceptual se utiliza en diferentes contextos para abstraer y representar diferentes componentes o entidades con sus atributos, interrelaciones entre dichos componentes y en general diversas situaciones presentes en sistemas o fenómenos de la vida real, lo que facilita su estudio y comprensión. Así mismo, son una herramienta para plantear e incidir sobre la realidad evaluada, para incluso transformarla [29].

El modelo propuesto en esta investigación busca ser la base para la construcción de una herramienta software que permita definir índices de riesgo de presuntas situaciones corruptas en convocatorias de contratación pública mediante la integración y comunicación de varios **servicios especializados**, los cuales, en su mayoría, están diseñados para abordar una situación específica donde hay probabilidad de corrupción, basado en los datos y documentos asociados a los procesos de contratación estatal en Colombia presentados por los proponentes y otros adquiridos de diversas fuentes de información como por ejemplo las cámaras de comercio y/o la registraduría nacional del estado civil. Estos servicios especializados son el corazón (core) principal del modelo y se encargan de detectar los diferentes probables delitos o anomalías relacionadas con corrupción en los procesos de contratación pública en Colombia. Los resultados arrojados por estos servicios brindan la información necesaria para determinar el índice de riesgo que presenta cada una de las convocatorias evaluadas en el modelo, con respecto a algún delito. Los servicios implementados e integrados en el modelo son:

**Detección de similitud entre documentos:** Servicio encargado de almacenar e indexar las propuestas técnicas asociados a una convocatoria mediante el procesamiento de texto por frases y términos, haciendo uso de técnicas de procesamiento del lenguaje natural y de recuperación de información, con el objetivo de identificar la similitud que existe entre una nueva propuesta técnica y una colección de propuestas previamente presentadas y almacenadas, mediante la comparación de los términos de cada una de las partes haciendo uso de técnicas como la similitud de cosenos, y la frecuencia de términos – frecuencia inversa del documento, logrando finalmente, identificar las frases que presentan similitud con las propuestas almacenadas.

**Detección de manipulación de ofertas:** Servicio encargado de, en primer lugar, almacenar el contenido de las propuestas económicas asociados a una convocatoria, mediante una extracción inteligente de datos. Con estos datos, el subproceso de análisis se encarga de realizar comparaciones de las diferentes

propuestas de una licitación, mediante el uso de medidas estadísticas como la varianza, el coeficiente de variación o la curtosis. Con el objetivo de encontrar anomalías que indiquen la probabilidad de presentarse el delito de manipulación de ofertas.

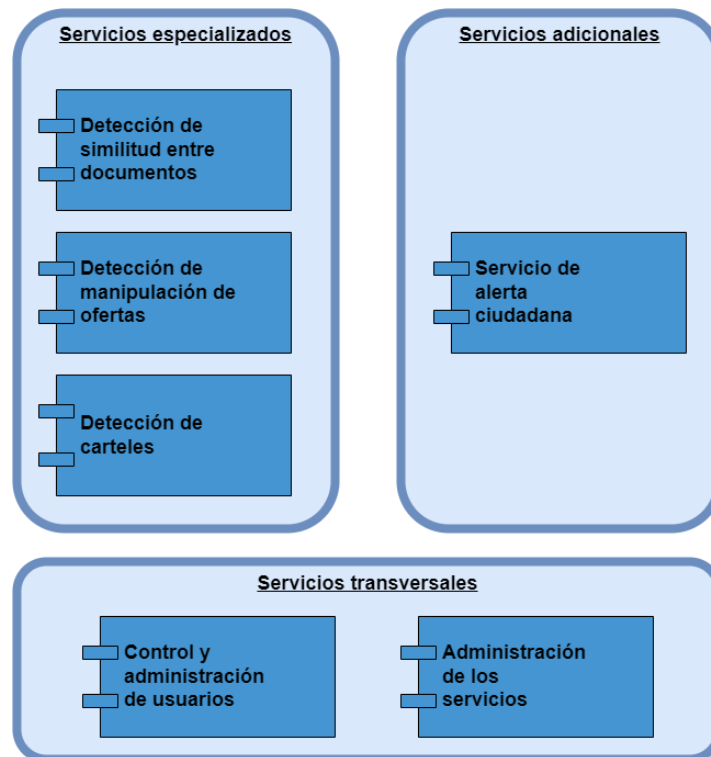
**Detección de carteles:** Servicio que, hace uso de los documentos recolectados en el proceso de cargue de una licitación, documentos referentes a información personal/empresarial del proponente, por ejemplo, el Registro único de proponentes (RUP), para así, poder crear una topología de red, que junto a una estrategia de análisis de imágenes permita la detección de patrones de comportamiento con el objetivo de alertar sobre la probable formación de carteles dentro de un proceso de licitación.

Por otro lado, se define un componente de **servicios adicionales** que pueden, sin necesidad de tratar un delito específico, ayudar a la identificación de situaciones anómalas en un proceso licitatorio. Este componente del modelo, entonces, alojará aquellos servicios adicionales que, tiene como objetivo generar una alerta temprana sobre las convocatorias existentes, sin necesariamente realizar análisis sobre la información existente, así, se podrían entender estos servicios, como servicios de índole social, con aplicaciones de las tecnologías de la información. Con base en una revisión de la literatura, y el análisis del impacto que el servicio de **alerta ciudadana** [30] ha tenido en propuestas previas, una adaptación de este fue incorporado en el modelo. En este modelo el servicio, plantea una recolección de información referente a quejas de situaciones anómalas dentro de un proceso de licitación, por medio de un servicio de denuncia ciudadana accesible al público en general. El servicio permite, además, de recolectar las denuncias hacerles su respectivo seguimiento, gestión y socialización de la investigación realizada por medio de reportes. Este servicio si bien no se enfoca en un delito como tal, puede permitir la detección de probables situaciones corruptas que los anteriores servicios podrían pasar por alto, dado que estos se enfocan detalladamente en un servicio específico como tal.

Además de los componentes de los servicios especializados y adicionales, se cuenta con un componente de **servicios transversales** que coadyuvan con el objetivo general del modelo, brindando características y funcionalidades vitales para la transformación del modelo en una herramienta software integrada de apoyo a las entidades públicas colombianas. Los servicios transversales implementados e integrados en el modelo son:

1. **Administración, autenticación y autorización de usuarios:** Servicio encargado del control, gestión de acceso, administración y autorización de usuarios y los permisos de estos sobre los componentes y servicios del modelo.
2. **Administración de los servicios:** Servicio encargado de gestionar las variables, atributos y características administrativas de los diferentes servicios del modelo.

En resumen, el modelo consta de 5 componentes principales, 1) los servicios especializados que buscan identificar situaciones específicas de probable corrupción, 2) los servicios transversales que soportan la transformación del modelo en una herramienta software, 3) servicios adicionales donde se abordan situaciones generales de probable corrupción, 4) las relaciones explícitas entre los servicios y 5) la salida global del modelo. En la **Figura 2** se muestra un diagrama donde se logra apreciar a un alto nivel los principales componentes del modelo. Este diagrama sigue una arquitectura orientada a eventos y delega las funcionalidades a través del concepto de microservicio.



**Figura 2.** Diagrama macro de componentes del modelo

### 3.1 Servicio de detección de similitud en propuestas técnicas

El plagio es definido por la Real Academia Española de la Lengua (RAE) de manera formal como la acción de copiar en lo sustancial obras ajenas, dándolas como propias. Esta acción, que es un problema que ha existido desde los albores de la escritura, es una práctica que en Colombia es reconocida como un delito por la legislación nacional en la Ley 1032 de 2006.

El delito de plagio se puede presentar de manera premeditada (intencionalmente) o de forma inconsciente (desconocimiento del concepto o de las formas para evitarlo), así mismo existen varias formas de llevar a cabo este delito [31], las cuales se mencionan a continuación:

- **Autoplagio:** Se presenta cuando un autor redacta nuevamente (de manera exacta o parafraseada) una obra que él mismo había realizado previamente.
- **Falsa autoría:** Se da una vez que se incluye a una persona como uno de los autores de una obra sin que este haya contribuido con la realización de esta.
- **Envío doble:** Se presenta al enviar una misma obra a dos eventos (conferencias, revistas, etc.) con el objetivo de, por ejemplo, obtener una mayor probabilidad de aceptación.
- **Robo de material:** Se da al copiar, sin autorización, contenido perteneciente a otro autor (es una de las formas más comunes y severas).

Además, el plagio puede presentarse o bien de tipo servil, donde el plagiador, se apropia textualmente de forma total o parcial de una obra, sin introducir modificaciones con respecto a la original, o de tipo inteligente, donde el plagiario, intenta “disimular” el delito (mediante el parafraseo, uso de sinónimos, etc.). Este último tipo es una modalidad más difícil de reconocer.

El delito de plagio afecta directamente al autor(es) de la obra que está siendo copiada afectando los derechos morales de su obra, así mismo el plagio atenta contra el interés público dado que, con esta práctica, se está engañando al “consumidor” final, haciéndole creer que un producto es único, cuando realmente es una copia del original. Por esto, es vital que este delito esté formalmente tipificado y legislado en la constitución política de un país, y que el mismo se haga cumplir correctamente. El delito de plagio, además, puede constituirse, dado el caso, en un hecho que atenta contra la administración pública, razón por la cual, en Colombia fue tipificado por el Estatuto Anticorrupción en la Ley 1474 del 2011 ***“Por la cual se dictan normas orientadas a fortalecer los mecanismos de prevención, investigación y sanción de actos de corrupción y la efectividad del control de la gestión pública.”***

Conceptualmente, la tarea de detectar plagio entre documentos se basa en la idea de localizar las partes de un documento que muestren índices de potencial plagio con uno o más documentos esto basado en algún cálculo aritmético [32]. A manera técnica, en la literatura, existen dos enfoques generales para la detección de plagio, siendo el enfoque extrínseco, el más usado para llevar a cabo esta tarea. El enfoque extrínseco consiste en comparar un documento al cuál se le desea conocer su probabilidad de plagio, contra una colección de documentos, previamente almacenados, que se supone son originales. Luego de llevar a cabo el análisis, se recuperan todos los documentos con los cuales se presentan similitudes por encima de un umbral previamente establecido [33], [34].

El plagio es un delito, que se puede presentar en cualquier ámbito, razón por la cual se han venido desarrollado herramientas para la detección de este. Las aplicaciones de detección de plagio deben abordar la problemática de la detección de similitud contemplando el rendimiento computacional de los cálculos que esto conlleve, es decir, hallar un equilibrio entre encontrar fuentes de plagio con costos computacionales razonables.

La gran mayoría, las aplicaciones anti-plagios desarrolladas están soportados en la web, aprovechando las conocidas ventajas que esto conlleva. La estrategia que estas aplicaciones utilizan para la señalización de las partes de un documento que presentan similitud con otra, es mediante el resaltado de dichas frases, apuntando a la fuente original de donde se pudo haber obtenido la información, dejando a manos del usuario la comprensión de cómo se produjo el probable plagio [35].

En la literatura referente al área se pueden encontrar diferentes estrategias, como en 2013 [36] donde recalcan la importancia de la implementación de estrategias automáticas para la detección de plagio, en la investigación implementan un algoritmo computacional basado en técnicas de procesamiento de lenguaje natural (Natural Language Processing, NLP). El algoritmo propuesto fue evaluado en una fuente de datos académicos, obteniendo resultados destacables.

Otra estrategia utilizada para la detección de plagio automática es el método basado en grafos propuesto en 2016 [37] que opera sobre la sintaxis y semántica de los contenidos, este enfoque logró resultados comparables con metodologías de NLP, pero con un costo computacional mucho más elevado.

En 2015 [38] los autores elogian el esfuerzo invertido para mejorar la detección de plagio basada en texto, sin embargo, señalan que hay una carencia de "técnicas capaces de identificar figuras, tablas, ecuaciones y documentos o imágenes escaneados". Razón por la cual en 2019 [39] los autores mencionan que si bien las aplicaciones actuales de detección de plagio encuentran de manera confiable casos de texto copiado (una de las maneras de plagio más usadas) y moderadamente redactado, estos presentan dificultad al detectar plagio oculto, como traducciones y reutilización de ideas no textuales, por lo que proponen una aplicación de detección que combina evaluaciones de similitud de contenido matemático y citas académicas. En la investigación utilizan el motor indexador Elasticsearch, resaltando los amplios beneficios que este otorga, en especial, a la hora de recuperar los documentos candidatos, haciendo uso de la medida TF/IDF (frecuencia de término – frecuencia inversa del documento) con un enfoque de recuperación binario. Sin embargo, la mayor parte de las investigaciones que atacan esta problemática todavía la abordan únicamente basado en texto.

El uso del motor de indexación Elasticsearch para realizar la detección de similitud entre documentos, aplicando estrategias no textuales es un enfoque que ha sido trabajado en diversas investigaciones entre 2016 y 2018 [40]–[42], mediante la implementación de una arquitectura basada en APIs, obteniendo resultados destacables y con costos computacionales reducidos.

Por otro lado, se destacan algunas investigaciones donde se han implementado estrategias que buscan analizar elementos no textuales dentro de los documentos para la identificación de plagio, demostrando el beneficio que esto conlleva para la tarea final, tal es el caso de [43] (2014) que, con el uso de análisis de patrones de citas en el texto, logran tasas de detección más altas que los enfoques léxicos en el ámbito académico, con un costo computacional moderado. En 2017 [44] se incluyó,

además, el análisis de imágenes y expresiones matemáticas obteniendo resultados que confirman que los métodos de detección no textuales mejoran significativamente las capacidades de detección. Esto demuestra que la tendencia de incluir métodos de análisis de contenido no textual es un componente prometedor en futuros enfoques de detección de similitud entre documentos.

En 2019 [45], luego de una revisión sistemática de la literatura de aplicaciones de detección de similitud en el ámbito académico, los autores destacan que existe una tendencia sobre los enfoques que integran diferentes métodos de detección con la ayuda del aprendizaje automático. De acuerdo con esta observación, los autores recalcan el gran potencial que estos enfoques tienen para la mejora de los métodos de detección de plagio en la integración de enfoques de análisis no textual (soluciones multimodales que incluyen texto e imágenes, por ejemplo). Sin embargo, mencionan que hay una falta de evaluaciones del desempeño de las aplicaciones de detección de plagio existentes a la fecha de la publicación (2020). Finalmente, los autores recalcan que los mecanismos utilizados a la fecha sólo resuelven el problema de la detección de plagio a corto plazo, por lo que invitan a invertir más en programas o mecanismos tempranos para su prevención, logrando así a largo plazo disminuir cada vez más la incidencia de esta problemática.

En resumen, luego de una revisión de la literatura, se aprecia cómo, diversas investigaciones han tratado de abordar el problema de la detección de plagio, utilizando diferentes enfoques, cada vez más complejos, con el objetivo de atacar el plagio “oculto”, sin embargo, la forma en cómo se realiza este proceso en la inmensa mayoría de aplicaciones estudiadas es similar, iniciando, o bien con un cargue de datos (fuente de información) o asumiendo la existencia del mismo, luego, la realización del análisis de un documento, basado en estrategias que varían desde cálculos aritméticos, implementación de técnicas de inteligencia artificial, entre otras. Terminando con la entrega de un reporte con la información al usuario final, señalando los lugares donde un documento es similar a otro. Con ello se puede concluir que una de las partes principales de un proceso de análisis de similitud o plagio entre documentos, es la necesidad de una fuente de información, desde donde se realizarán las comparaciones, de no existir dicha fuente, el primer paso debe ser, entonces, el almacenamiento sistemático de la información, con la ayuda de un motor de indexación que facilite la tarea, seguido de esto se debe escoger la estrategia de análisis y entrega de resultados que será objeto de modelado, diseño, desarrollo y prueba antes de su salida a producción.

Es necesario destacar que, en la literatura, la inmensa mayoría de investigaciones dedicadas a la detección de plagio lo hacen enfocando el problema en un ámbito académico y son nulas investigaciones dedicadas a abordar el problema en los documentos asociados al proceso de contratación, lo que se presenta como un aporte de este trabajo y una oportunidad de investigación.

En esta investigación se presenta un servicio especializado para la detección de similitudes entre documentos asociados a un proceso de contratación pública, específicamente los documentos que presentan las “**propuestas técnicas**”. Esto

mediante el uso de técnicas de procesamiento del lenguaje natural como, recuperación de la información. Así mismo el servicio cuenta con una funcionalidad de almacenamiento sistemático de información que podrá ser usado por posteriores servicios especializados para atacar otro tipo de delitos relacionados con el proceso de contratación.

### **3.1.1 Datos de entrada requeridos**

El objetivo de este servicio consiste en identificar las probables similitudes que se presentan entre una nueva propuesta técnica y una fuente de información que se irá construyendo a partir de las propuestas técnicas que sean analizados en un proceso de licitación actual o previo. El servicio se ha modelado con el objetivo de generar alertas sobre una propuesta técnica específica, señalando las partes de esta que presentan similitudes por encima de un umbral previamente definido. Siendo el usuario final quien debe verificar si el reporte entregado conlleva a marcar la propuesta técnica, unas secciones de esta o unas frases como plagio.

Para lo anterior se debe partir, entonces, de la existencia de la propuesta técnica, el archivo donde un proponente expone las condiciones de carácter técnico del bien o servicio a ofrecer de acuerdo con lo previamente exigido por la entidad. Esta propuesta técnica en el marco de una licitación es obligatoria, enmarcado en el artículo 7, de la Ley 1150 de 2007. Y según el numeral 8º del artículo 30 y el numeral 15 del artículo 25 de la Ley 80 de 1993. Esta propuesta técnica no podrá complementarse, adicionarse, modificarse o mejorarse cuando se cierre la convocatoria (licitación).

Entonces, un proponente que desee participar dentro de un proceso de licitación, que tiene como objetivo la adquisición de algún producto o servicio debe entregar, cumpliendo con el pliego de condiciones de una licitación, una propuesta técnica donde expresa con claridad las condiciones técnicas a desarrollar (además de cumplir con toda la documentación del caso). Esta propuesta técnica, es la fuente de información que el servicio de detección de similitud usará para, 1) retroalimentar la fuente de propuestas técnicas desde donde se comparará una nueva y a su vez 2) es el documento al que se le realizará un análisis de similitud frente a las propuestas técnicas previamente recibidas y almacenadas.

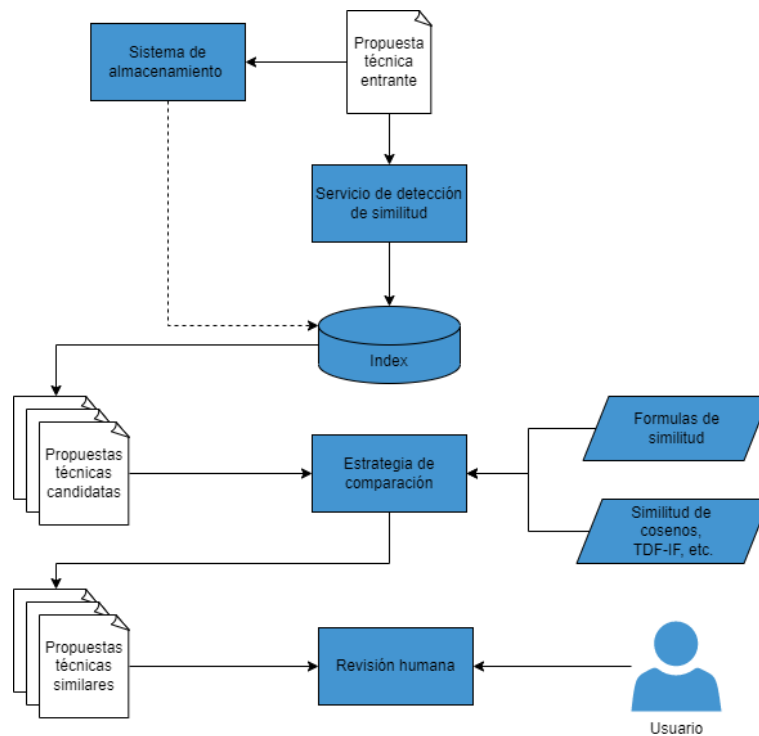
La estructura o presentación de la propuesta técnica y su información puede variar, dependiendo de lo establecido por cada entidad para cada licitación, sin embargo, por lo general, y como propósito del correcto funcionamiento del servicio, se parte de la idea de que la propuesta técnica se presenta como un archivo en formatos conocidos (extensiones .pdf, .doc, y .docx). Ahora bien, existe la posibilidad (que no será abordado en esta investigación) de que el archivo se encuentre escaneado y que se adjunte en formato de tipo imagen (.jpg, .png, etc.). Para abordar este problema, se debe primero contar con un servicio de extracción de texto en imágenes.



### 3.1.2 Procedimiento

El servicio de detección de plagio propuesto en esta investigación consiste en un analizador de texto para la identificación de similitudes de contenido potencialmente sospechosas. El público objetivo del servicio, al igual que el modelo como tal, son los tomadores de decisiones sobre los procesos de licitación en una entidad.

El enfoque propuesto para la implementación del servicio se puede detallar en la **Figura 3**. El enfoque sigue los principios de diseños de aplicaciones anti-plagios de la literatura, basándose en las siguientes etapas: recuperación de propuestas técnicas candidatas, estrategia de comparaciones detalladas y la inspección humana [46]. Además de ello, dada la necesidad de construir y retroalimentar la base de datos fuente de información, el servicio realiza una primera etapa de almacenamiento sistemático de propuestas, es válido recalcar que esta etapa se debe realizar en un primer momento, con el objetivo de poblar la fuente de datos y a su vez, es un proceso iterativo e incremental, dado que se realizará en repetidas ocasiones, alimentando así, cada vez más, la fuente de información.



**Figura 3.** Descripción general del proceso de detección de similitud

Para el almacenamiento de las propuestas técnicas, se debe partir de que el mismo se encuentra en algún formato específico, su contenido debe ser extraído ya sea para indexarlo o analizarlo, y en este proceso en algunas circunstancias se requiere dividir el contenido en sus frases correspondientes. Una ventaja a la hora del desarrollo de estas funcionalidades es que actualmente hay librerías de código abierto y de uso libre que realizan estas tareas, lo cual evita su desarrollo desde cero. Así mismo, el archivo original de la propuesta técnica se debe almacenar en

el sistema de archivos y relacionarlo en la base de datos para llevar un apropiado control de los posteriores análisis. El almacenamiento en sistemas de archivo es una estrategia que se usa en lugar del almacenado directo en base de datos para disminuir costos (más costoso el almacenamiento en bases de datos que en sistemas de archivos). De esta forma, en la base de datos solo se almacena la ruta o referencia donde está guardado el archivo en disco.

Una vez extraído el contenido de la propuesta técnica, el primer paso del procesamiento del archivo corresponde a la indexación, de manera que pueda ser utilizado posteriormente para el análisis de similitud, para ello se hace uso de los índices invertidos, gracias a las amplias ventajas que ofrecen a la hora de realizar búsquedas complejas sobre los datos [47], tema de gran importancia para esta investigación. A parte de esta estrategia, se hace necesario contar con una base de datos que almacena el archivo y lo relaciona con información externa del mismo, evitando así sobrecargar el indexador con datos que no son representativos para la posterior búsqueda. Es válido mencionar que, para una correcta indexación, el índice debe ser configurado, modificando las operaciones (filtros, reducción a la raíz léxica, remoción de palabras vacías, entre otras) que se debe realizar en el momento de la indexación, esto ayuda a que el proceso de búsqueda sea más eficiente, permitiendo búsquedas avanzadas de mayor complejidad.

Hoy en día la literatura experta reporta algunas herramientas de libre uso que permiten realizar almacenamiento y búsqueda en índices invertidos. En la presente investigación se hizo uso de **Elasticsearch**, un motor de búsqueda basado en **Lucene** que a su vez es una API de código abierto dirigido a la tarea de recuperación de información. **Lucene** se encuentra desarrollado en Java, pero su API se encuentra disponible en una amplia variedad de lenguajes de programación, ahora bien, **Elasticsearch** que tiene su core en **Lucene** para las tareas de indexación y búsqueda, ofrece sus funcionalidades de una manera más práctica, y ofrece además la implementación de una aplicación distribuida mediante nodos que le permite lograr una mayor disponibilidad, el cargue masivo de documentos y herramientas de búsqueda y análisis.

Por otro lado, la recuperación de propuestas técnicas candidatas y la estrategia de comparaciones detalladas de estas, es el core del servicio de similitudes entre propuestas técnicas. Para este caso en específico, se contempla, inicialmente una estrategia para tener en cuenta las características del documento, convocatoria y tipo de análisis a desarrollar, luego de esto, se hace uso del indexador para el envío de la petición de búsqueda, aplicando filtros y configuraciones específicas de la API de **Elasticsearch**. Un aspecto por tener en cuenta en esta tarea es que el análisis se genera basado en las propuestas técnicas que se encuentren indexados en el momento, entonces, si luego de indexar nuevas propuestas técnicas se desea que una propuesta técnica en específico tenga en cuenta estos nuevos registros, se debe realizar un nuevo análisis de similitud. El servicio, entonces debe contar con un histórico de análisis realizados para cada propuesta técnica.

Finalmente, para la tarea de inspección humana de los análisis resultantes, se debe tener en cuenta, inicialmente, que los mismos toman un tiempo para su realización, esto debido a: 1) la gran cantidad de palabras o frases de la propuesta técnica que se va a analizar y 2) la gran cantidad de propuestas técnicas con las que se debe realizar la comparación. Por lo anterior se hace necesario implementar una estrategia de entrega asincrónica del reporte, permitiendo así que el usuario continúe trabajando en la aplicación sin la necesidad de esperar a que la tarea de análisis termine. Una vez la tarea concluye, se informa al usuario de la disponibilidad del reporte de análisis a través de una notificación.

Para esta tarea se debe contemplar, primero el almacenamiento de la información de los análisis en un motor de base de datos no relacional, esto debido a la estructura del reporte y con el objetivo de no tener que repetir el proceso de análisis cada vez que se desee visualizar, sino más bien hacer uso de la información previamente almacenada. Y finalmente una estrategia para la entrega y visualización del reporte. Además, una estrategia para la configuración de los análisis generados también debe ser implementada, con el objetivo de brindarle al usuario final una fácil administración de estos, como por ejemplo poder etiquetar aquellas frases similares que no deben contar a la hora de cuantificar la similitud del documento (una lista de “frases comunes” que no deben contar como plagio).

Con el objetivo de entender las principales acciones e interacciones que involucran el servicio de detección de similitudes, en la **Figura 4** se presenta un diagrama de procesos de negocio que modela el funcionamiento de este. En la figura se logra apreciar, como el proceso inicia con la petición por parte del usuario, partiendo del cargue de propuestas técnicas, de la cual se dividen dos subprocesos, uno para el almacenamiento de archivos y el otro para la extracción de datos, de donde se despliegan las dos siguientes tareas del usuario, la indexación de propuestas técnicas y el análisis de similitud, las dos se comunican con un servicio de indexación para el desarrollo de sus respectivas tareas. Una vez realizado el cargue de información, el proceso finaliza con la entrega bien sea, de la referencia de la indexación creada, o el propio análisis de similitud, en ambos casos se genera una notificación del estado final de la tarea al usuario.

### 3.1.3 Salida

El objetivo de este servicio, al igual que los demás servicios del modelo propuesto, más que juzgar una propuesta, busca alertar al tomador de decisiones sobre irregularidades que se puedan presentar en los documentos. Para este servicio en específico se maneja un índice de similitud sobre el documento analizado, el cuál es un valor porcentual que indica, la semejanza de una propuesta técnica frente a otras almacenadas en la fuente de datos, esto basado en la similitud que exista entre las frases de las diferentes propuestas. Este porcentaje se entrega de manera inicial al tomador de decisiones, pero puede ser alterado, dado que el usuario puede realizar ajustes en las “frases comunes” que identifique en la propuesta.

La salida de este servicio puede ser visualizada en un reporte de similitud, similar a los utilizados por las aplicaciones encontradas en la literatura que abordan la problemática de la detección de plagio entre documentos.

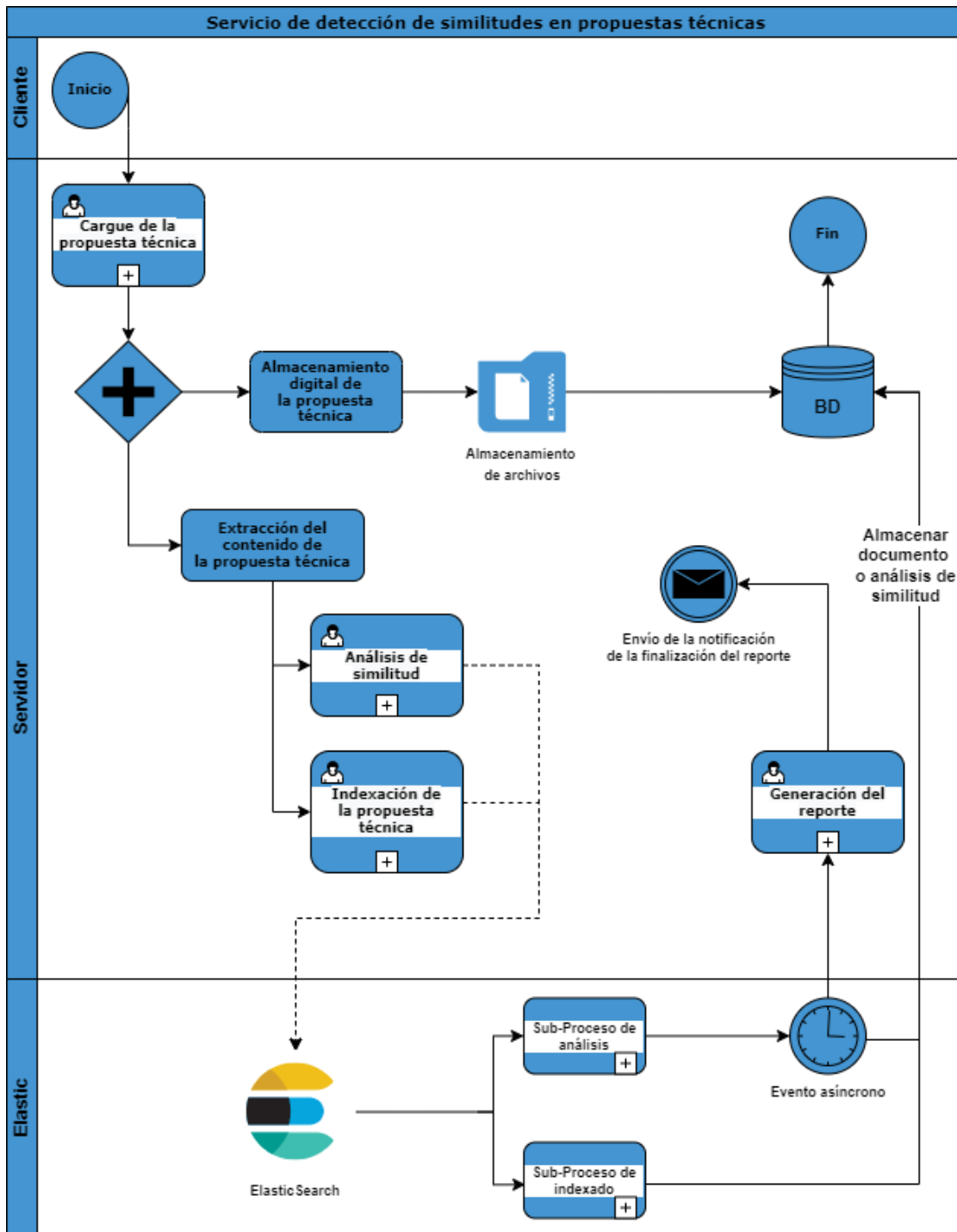


Figura 4. P.N. del módulo de detección de similitudes

### 3.2 Servicio de detección de probable manipulación de ofertas

La manipulación de ofertas, también conocida como amaño de ofertas, manipulación de licitaciones o colusión, es un delito enmarcado en los casos de fraude de corrupción contractual. Este delito se presenta cuando en un proceso de licitación, los proponentes confabulan entre sí para la elección del ganador del proceso, mediante el aumento o reducción de los precios y la calidad de los productos o servicios ofertados. Lo ideal para una entidad es obtener un equilibrio entre el costo y la calidad de los productos o servicios, para ello se recurre a un proceso de licitación competitivo, pero esto solo es efectivo si los competidores contienden con honestidad y transparencia. La manipulación de ofertas resulta particularmente dañina para un Estado cuando se ven afectados los rubros y adquisiciones públicas.

Según algunas investigaciones, en un proceso de licitación pública el delito de manipulación de ofertas se puede presentar de diferentes formas, aunque todas ellas pretenden lo mismo, a saber: 1) Impulsar un ganador previamente escogido por parte de todos o un segmento de los ofertantes, o 2) Elevar el monto de la oferta ganadora con el propósito de aumentar las utilidades del proponente ganador. Las principales modalidades en las que se puede presentar este delito están enmarcadas en cuatro tipos [48], los cuales se mencionan a continuación:

- **Ofertas de resguardo:** Este tipo de ofertas, (conocidas también como complementarias o simbólicas) son las más comúnmente encontradas en este tipo de delito. Esta se presenta cuando los proponentes de una licitación acuerdan licitar ofertas que contempla por lo menos uno de los siguientes elementos: (1) presentar una oferta más alta que la del ganador previamente designado, (2) presentar una oferta extremadamente elevada, la cual no es competitiva, o (3) presentar ofertas que contienen términos inaceptables por parte de la entidad compradora. Esta modalidad está diseñada para “resguardar” al ganador, aparentando una competencia legítima.
- **Supresión de ofertas:** Esta modalidad consiste en el establecimiento de acuerdos entre los licitantes para garantizar que uno o varios de ellos se abstengan de presentar o retiren una oferta presentada con anterioridad.
- **Rotaciones de oferta:** En esta modalidad los licitantes acuerdan “tomar turnos para ganar”, de esta forma, quien está de turno presenta la mejor oferta para la licitación.
- **Asignaciones de mercado:** En esta modalidad los proponentes se reparten bien sea por zonas de mercados o zonas geográficas evitando así la competencia entre ellos.

En 2014 [48] recalcan que estas modalidades no son mutuamente excluyentes por lo que se pueden presentar más de un tipo en un mismo proceso de licitación. Estos tipos pueden generar una serie de patrones que pueden ser detectados por parte de los funcionarios encargados de las entidades licitantes.

Varias investigaciones se han llevado a cabo con el propósito de identificar, detectar o predecir el delito de manipulación de ofertas en un proceso de licitación, estas se pueden agrupar en: 1) Métodos econométricos, 2) Métodos sociales, y 2) Métodos estadísticos y computacionales. A continuación, se presentan las más relevantes en cada grupo ordenadas cronológicamente.

### **Métodos econométricos:**

En 1993 [8] se recalca que en general, encontrar un único procedimiento de prueba para detectar la manipulación de ofertas es un objetivo muy difícil de llevar a cabo (por lo menos en el momento de la publicación de la investigación). Proponen unos procedimientos econométricos de prueba, diseñados con el objetivo de detectar la presencia de aparejos de licitación en subastas de adquisiciones en el área de la construcción. Los datos arrojados en este estudio mostraron que, aparentemente la manipulación de ofertas no tomó la forma de un esquema de rotación de ofertas. Sin embargo, varios licitantes ofertaron en la mayoría de las licitaciones en donde solo uno era un licitante válido y los demás presentaban ofertas falsas más elevadas.

En 2013 [9] se propone un método econométrico de detección de manipulación de ofertas que permite revelar carteles en subastas de adquisición sin ningún conocimiento previo de la estructura del mercado. El método implementado consiste en dos pasos, a saber: 1) Se tiene el supuesto de que, si existe una manipulación de ofertas, es probable que se generen precios relativamente altos, que normalmente se concentran cerca del valor estimado del contrato. En el primer paso, todas las subastas se dividen en dos grupos utilizando el precio relativo (la oferta del ganador dividida por el precio estimado del contrato) y 2) El segundo paso consiste en encontrar diferentes patrones de comportamiento entre los grupos.

### **Métodos sociales:**

En 2018 [7] se introduce un modelo derivado de las teorías criminológicas (como el triángulo de fraude de Cressey), compuesto por mecanismos de gobierno corporativo, controles internos y análisis de banderas rojas, que las organizaciones pueden aplicar para prevenir la corrupción. El modelo no es planteado en términos matemáticos sino más bien sociales, por lo que llevarlo a un marco tecnológico es una tarea compleja, sin embargo, el documento ofrece algunos aspectos interesantes para tener en cuenta en el proceso de detección de la corrupción, como la presencia de patrones de compras anormales que servirían para la creación de reglas de comportamiento, en el delito de manipulación de ofertas.

### **Métodos estadísticos y computacionales:**

En 2001 [28] con el objetivo de lograr establecer diferencias entre propuestas económicas corruptas y competitivas, se propone un algoritmo computacional que puede usarse en la detección empírica de la colusión, logrando, por ejemplo, establecer algunas reglas como el hecho de que las ofertas competitivas siempre satisfacen una propiedad que es llamada en la investigación, “intercambiabilidad”,

haciendo referencia al caso en que los costos del licitante 1 son cambiados por los costos del licitante 2, manteniendo todo lo demás fijo, y aun así la oferta funciona de forma similar. A diferencia de lo anterior, los modelos colusorios no siempre satisfacen esta propiedad. Los autores también mencionan que los algoritmos computacionales desarrollados en este trabajo pueden usarse para la evaluación post mortem de los daños causados por la manipulación de ofertas.

En 2003 [12] se propone un modelo de técnicas estadísticas basado en el teorema de Bayes. Similar a estudios previos, en esta investigación los autores proponen un modelo que establece un conjunto de condiciones, tales como independencia condicional e intercambiabilidad, que son necesarias y suficientes para que una distribución de ofertas sea calificada como una licitación competitiva. Así mismo, los autores discuten cómo se pueden probar estas condiciones argumentando que la decisión entre licitación competitiva y colusoria puede formularse naturalmente como un problema de decisión estadística, destacando además, la importancia de la implementación de herramientas computacionales que puedan ser utilizadas para detectar la manipulación de ofertas, y que esto, a su vez, pueda ser un elemento disuasorio para la colusión futura y, por lo tanto, reducir los precios y mejorar la eficiencia en algunas industrias.

En 2012 [11] se propone un método paso a paso, que hace uso de niveles de correlación para detectar las ofertas colusorias mediante el análisis de los datos históricos de ofertas proporcionados por el propietario de una construcción. El método de detección propuesto se basa en un modelo de regresión e implica identificar a los probables postores del cártel utilizando dos pruebas llamadas, prueba residual y prueba de estabilidad de la estructura de costos. En el artículo se recalcan algunos probables signos de colusión identificados por el Departamento de Justicia de USA quien observó algunos patrones de comportamiento de licitación que contrastan con el comportamiento competitivo del mercado. Algunos de esos patrones fueron:

- El mismo grupo de postores presentan ofertas y los postores del grupo se turnan para ser el postor más bajo.
- Un licitante ofrece ofertas significativamente más altas en algunas ofertas que en otras, sin diferencias de costo perceptibles que explique el aumento del precio.
- Los precios de oferta bajan cada vez que un postor nuevo o poco frecuente presenta una oferta.
- Hay menos postores en comparación con las instancias de licitación normales.
- El licitante a quien se le adjudica el contrato ha trabajado para los competidores que presentaron ofertas fallidas en el mismo proyecto.
- Los postores presentan ofertas idénticas donde (1) los precios permanecen idénticos durante largos períodos de tiempo, (2) los precios anteriores eran diferentes y (3) el aumento del precio no parece estar relacionado con el aumento de los costos.

Estos patrones encajan en las modalidades anteriormente mencionadas, en las cuales se puede presentar este delito. Una limitación importante de este estudio es que el modelo propuesto no puede detectar las ofertas colusorias en el momento en que sucede; es capaz de detectar sospechas de colusión en el pasado y solo si el analista tiene acceso a información histórica detallada sobre un conjunto de variables independientes.

En 2017 [49] se señaló que la industria de la construcción es una de las áreas más vulnerables en la ocurrencia de este delito debido a los altos montos que se manejan en los contratos de este tipo. Así mismo, mencionan la importancia de desarrollar nuevas herramientas, utilizando técnicas y recursos que antes no estaban disponibles para abordar esta problemática, resaltando la utilización de un esquema de licitaciones electrónicas para contratos públicos, en donde la entidad ordenante podría impartir algunas directrices sobre los documentos a presentar siguiendo un modelo previamente establecido, logrando así el almacenamiento de datos sobre cada licitación a lo largo de los años, permitiendo con ello el análisis de dicha información que serviría para aplicar estrategias tanto como para abordar este delito o nuevos tipos de delitos.

En una revisión de la literatura sobre investigaciones relacionadas con la detección, identificación o predicción del delito de manipulación de ofertas en el proceso de licitación, en 2018 [50] destacaron que la detección de este delito es una tarea difícil que requiere un amplio conocimiento interdisciplinario, experiencia, selección de herramientas adecuadas y disponibilidad de datos seleccionados. Los autores también recalcan que mediante una base de datos debidamente diseñada y actualizada sistemáticamente podrían llevarse a cabo análisis mediante el uso de inteligencia artificial con el objetivo de detectar irregularidades, relaciones que indiquen la probabilidad de colusión, como por ejemplo la implementación de reglas de asociación que generalmente son difíciles de observar en una enorme base de datos. Así mismo, mencionan que la implementación de redes neuronales artificiales podría usarse para clasificar las licitaciones en diferentes grupos de riesgo de ocurrencia de colusión en la etapa de análisis preliminar de la licitación. Los autores terminaron concluyendo que la implementación de minería de datos sobre la amplia fuente de información que se puede formar podría permitir la recuperación de datos muy valiosos de las ofertas almacenadas, las cuales servirían para poder atacar además otros tipos de delitos.

También en 2018 [10] se propone un método estadístico para detectar la manipulación de ofertas mediante la aplicación de pantallas (índices estadísticos específicos) que se refuerzan mutuamente, sobre un conjunto de datos de licitaciones de adquisición de materiales de construcción para carreteras de Suiza. En este trabajo se destaca el hecho de que no se disponía de información previa sobre colusión. El método de detección es particularmente adecuado para abordar el problema de la colusión parcial, es decir, la colusión que no involucra a todas las empresas y/o todos los contratos en un conjunto de datos específico. En general, este enfoque equivale a un conjunto de pruebas que se refuerzan mutuamente para identificar una probable colusión entre subconjuntos de empresas, arrojando un



índice de probabilidad de que una oferta presentada en un proceso de licitación sea colusoria.

En 2019 [18] se combinaron técnicas de aprendizaje automático con pantallas estadísticas calculadas a partir de la distribución de ofertas en licitaciones dentro del sector de la construcción en Suiza. Este trabajo buscó predecir la colusión a través de carteles de manipulación de ofertas. El modelo creado otorga una clasificación binaria en la cual se informa si una oferta es marcada como probable colusoria, clasificando como 1 si la probabilidad de colusión de la oferta es superior a 0,5 y 0 de lo contrario. Los resultados sugieren que la combinación de aprendizaje automático y pantallas estadísticas es una herramienta poderosa para detectar la manipulación de ofertas, dado que luego de aplicar el método de validación cruzada (cross validation), como mecanismo de evaluación del rendimiento de este enfoque se encontró que el mismo logró clasificar correctamente más del 84% del total de los procesos de licitación como colusorios o no colusorios. Los métodos propuestos en este documento exhiben un rendimiento bueno, ya que la proporción de falsos positivos y negativos es significativamente baja. Además, los autores sugieren ajustar la regla para clasificar las ofertas como colusorias, prediciendo la colusión solo como 1 si la probabilidad de colusión predicha por el algoritmo excede un umbral específico más alto que el valor predeterminado de 0.5. Una de las principales conclusiones que logran los autores es que las dos pantallas más potentes en el modelo son el coeficiente de variación y la distancia normalizada. La distancia relativa normalizada se asocia positivamente con la probabilidad de colusión. Esto implica que la manipulación de ofertas reduce la varianza de las ofertas y conlleva una asimetría en la distribución de las ofertas. Finalmente, en el artículo se discute sobre las implicaciones políticas para las agencias que tienen como objetivo detectar carteles de manipulación de ofertas.

En 2021 [51] aunque recalcan la importancia de los trabajos realizados en [28], [30] y [31] mencionan que los documentos hacen referencia solo a una forma de manipulación de licitaciones: cárteles completos de manipulación de licitaciones, mencionando al mismo tiempo que la realidad es más compleja y, a menudo, en una misma licitación se pueden presentar proponentes competitivos y proponentes colusorios pertenecientes a un cartel. Los autores de esta investigación mencionan que las ofertas competitivas distorsionan las señales estadísticas producidas por los cárteles en la distribución de las ofertas de modo que los métodos de detección convencionales no reconocen el rastro de la manipulación de licitaciones, por lo que en esta investigación se propone un método de detección más robusto, mediante la combinación de pantallas estadísticas derivadas de la distribución de ofertas en una licitación con aprendizaje automático para predecir la probabilidad de colusión. Los resultados arrojados obtuvieron una tasa de clasificación destacada incluso en presencia de ofertas competitivas.

En resumen, se logra identificar como a lo largo de los años, diversas investigaciones han tratado de abordar el problema de la identificación del delito de manipulación de ofertas en procesos de licitación con diferentes metodologías, las cuales han ido evolucionando con el paso del tiempo, pasando de test

econométricos a técnicas estadísticas y últimamente, mezclando estas últimas con técnicas computacionales, aprovechando los amplios beneficios que se pueden obtener de esto. El futuro de las soluciones a esta problemática, entonces es el aprovechamiento que se le pueda dar a los grandes volúmenes de datos y documentos referente a las ofertas que se puedan ir almacenando por cada licitación realizada, esto claramente, necesita un modelo previamente definido, en donde, primero se garantice el almacenamiento sistemático de dicha información, dado que, como se ha mencionado en muchas investigaciones, la falta de información es el principal obstáculo para atacar este problema. En este sentido, se debe ahondar en la aplicación de técnicas de minería de datos, inteligencia artificial y demás herramientas de TI que permitan sacarle provecho a los datos con el objetivo de atacar este delito de manipulación de ofertas.

### **3.2.1 Normas vigentes**

Basado en el marco legal de la contratación pública en Colombia que se encuentra enmarcado en las leyes 80 de 1993 y 1150 de 2007, y el decreto 1510 de 2013, se establecen, entre otras cosas, los procedimientos para contratar y las modalidades de selección del contratista, los cuales se resumen a continuación.

1. En el marco de una licitación pública, el primer paso que debe dar una entidad consiste en identificar la necesidad de adquirir productos o servicios, para lo cual elabora un plan de compras, el cual debe ser aprobado por un ente regulador.
2. La entidad elabora un pliego de condiciones a tener en cuenta en el proceso, el cual debe ser publicado en el SECOP (plataforma en la cual las Entidades Estatales deben publicar los Documentos del Proceso, desde la planeación del contrato hasta su liquidación. También permite a las entidades estatales y al sector privado tener una comunicación abierta y reglada sobre los procesos de contratación) por un término mínimo de diez (10) días hábiles, así como la debida publicación del aviso de la licitación.
3. Todo aquel interesado en el proceso puede solicitar aclaraciones sobre el pliego de condiciones, dentro de los ya mencionados diez (10) días hábiles.
4. Se da apertura el inicio del proceso, considerando las observaciones realizadas en el periodo de aclaraciones.
5. Conforme al cronograma establecido en los términos de la licitación, los interesados envían las propuestas junto con la documentación establecida.
6. La entidad inicia la etapa de verificación de requisitos y evaluación de las propuestas que cumplen con lo establecido asignando puntajes a las ofertas.
7. El informe de evaluación se pone a disposición de los participantes, para su conocimiento público, con el objetivo de realizar observaciones al proceso.
8. La entidad pública analiza las observaciones y cita un proceso de audiencia pública, según lo establecido en el cronograma.
9. Una vez instalada la audiencia los proponentes pueden pronunciarse sobre circunstancias del proceso sin que ello implique una nueva etapa de “modificación de oferta”.
10. Si algún proponente cumple con lo previamente establecido, la entidad pública adjudica el contrato, de lo contrario se declara como desierto el proceso.

A continuación, se presenta un servicio especializado, basado en los datos y documentos recolectados con las propuestas económicas, en el marco de un proceso de licitación, el cual hace uso de técnicas de procesamiento del lenguaje natural y recuperación de la información con el objetivo de poder identificar si en una licitación existe la probabilidad de presentarse el delito de manipulación de ofertas.

### 3.2.2 Datos de entrada requeridos

Como se logró apreciar en la revisión de la literatura realizada, muchos de los estudios pretenden abordar la identificación de la manipulación de ofertas, identificando la presencia de grupos colusorios, mediante el histórico de propuestas realizadas. En esta investigación, el objetivo de este servicio es la evaluación de las propuestas económicas presentadas a una licitación abierta para encontrar probables anomalías en estas, referentes a los costos de los servicios y/o productos de las diferentes propuestas, con ello, el servicio busca generar alertas sobre la probable manipulación de las ofertas. Teniendo en cuenta que como base de todos los servicios se cuenta con el almacenamiento de documentos de propuestas anteriormente presentadas a diversas licitaciones, estas mismas pueden ser utilizadas para hacer estudios post mortem que permitan identificar diferentes anomalías en las licitaciones realizadas, con esto aprender y perfeccionar los servicios que se proponen en este trabajo, incluyendo por ejemplo la identificación de colusión basada en asignaciones de mercado o territorio.

Un proponente que desee participar dentro de un proceso de licitación abierto debe entregar, cumpliendo con el pliego de condiciones de una licitación, una propuesta económica donde se exprese con claridad las exigencias realizadas por la entidad pública (además de cumplir con toda la documentación requerida). Esta propuesta económica relaciona el o los productos o servicios con su descripción, cantidad y costo propuesto por el licitante. Este servicio parte de la existencia de este documento, la propuesta económica. La estructura o presentación de estos datos pueden variar, dependiendo de lo establecido por cada entidad para cada licitación, pero por lo general, y es lo que se establece para el correcto funcionamiento de este servicio, se maneja una tabla con la estructura que se presenta en la **Tabla 7**.

**Tabla 7.** Tabla de precios de la propuesta económica

Identificación	Nombre del producto/servicio	Descripción del producto/servicio	Descripción del producto/servicio ofertado	Valor unitario	Cantidad	Valor total
Identificación que denota y/o diferencia un producto o servicio. Puede ser un código alfanumérico	Nombre del producto/servicio que la entidad identifica que es necesario adquirir. Este nombre también lo establecen	Descripción detallada del producto/servicio solicitado por la entidad pública. En muchos de los casos este ítem es opcional, sin	Si el producto o servicio que presenta el oferente no es igual al solicitado en la licitación, en este espacio se describe. En	Valor monetario en el cual el participante de la convocatoria plasma su propuesta económica al producto/servi	Valor numérico que especifica la cantidad de productos/servicios solicitados. Este valor es publicado en el pliego de condiciones (términos de	Valor monetario resultante de la multiplicación entre el valor unitario y

co o un consecutivo o establecido en los términos de referencia (pliego de condiciones) de la licitación.	los términos de referencia (pliego de condiciones) de la licitación.	embargo, el mismo sirve de ayuda para la correcta definición del producto/servicio requerido. Este valor es publicado en el pliego de condiciones.	caso de productos, sirve para detallar las marcas y modelos específicos para lo cual siempre se debe anteceder a estas características la palabra apropiada (marca, modelo, entre otras).	cio. Debe ir en la unidad de medida solicitada en los términos de referencia de la licitación, por ejemplo, pesos colombianos.	referencia) de la licitación.	la cantidad.
---	--	--	---	--	-------------------------------	--------------

Para la implementación del servicio, el único campo que no es obligatorio es el relacionado con la columna 3 (Descripción del producto/servicio) y los identificadores deben ser exactamente los mismos que se suministran en el pliego de condiciones de la licitación. Por otro lado, los nombres de los productos y servicios deberían estar normalizados a nivel nacional, algo que está realizando actualmente el SECOP.

El formato en el cual se entregue esta tabla puede ser variado (hoja de cálculo, archivo PDF, documento de texto separado por comas u otro), siempre y cuando cumpla con la estructura anteriormente mencionada. Sin embargo si se deseara recolectar información de anteriores convocatorias, para un futuro uso, y buscando mejorar el servicio o para atacar otro delito, es necesario crear una estrategia para la extracción de los datos necesarios, esto dado que los documentos históricos donde se requiere extraer los datos muy seguramente no cumplen con el formato mencionado. Esta estrategia se menciona en el siguiente apartado.

### 3.2.3 Procedimiento

La **Figura 5** presenta un diagrama de procesos de negocio que modela el funcionamiento del presente servicio. En la figura se logra apreciar, como el proceso inicia con una petición realizada por un usuario. Parte del cargue de la propuesta económica de cada uno de los proponentes (tarea que se realiza previamente en un primer paso, en el momento del almacenamiento de la información que todo proponente debe entregar de acuerdo con el pliego de condiciones), luego se inicia el proceso de extracción inteligente de los datos (se explica en mayor detalle en la siguiente sección) necesarios para el análisis, este proceso además de extraer los datos los almacena. Con estos datos, el subproceso de análisis se encarga de realizar comparaciones de las diferentes propuestas de una licitación con el objetivo de encontrar anomalías que indiquen la probabilidad de presentarse el delito de manipulación de ofertas (se explica en mayor detalle en una sección posterior a la siguiente), los escenarios que debe tener en cuenta el servicio para desarrollar este análisis son:

1. Existencia de ofertas globales anormales (que sean muy elevadas o inferiores a las demás).
2. Existencia de manipulación y variación de ofertas individuales que pretendan aparentar una competencia leal al reunir grupos de ofertantes que manipulen los valores en productos/servicios específicos dentro de una convocatoria (ver detalles en la sección posterior a la siguiente).
3. Agrupamiento de ofertas que son similares en características y/o valores individuales o globales.

Finalmente, una vez realizado el análisis, el proceso culmina con la entrega del reporte, dentro de un motor de base de datos y la generación de una notificación del estado final de la tarea al usuario. Es preciso destacar, que previamente se realice la verificación del cumplimiento de los términos de referencia por parte de las ofertas, ya que, si esto no ocurre, estas deben ser directamente descartadas.

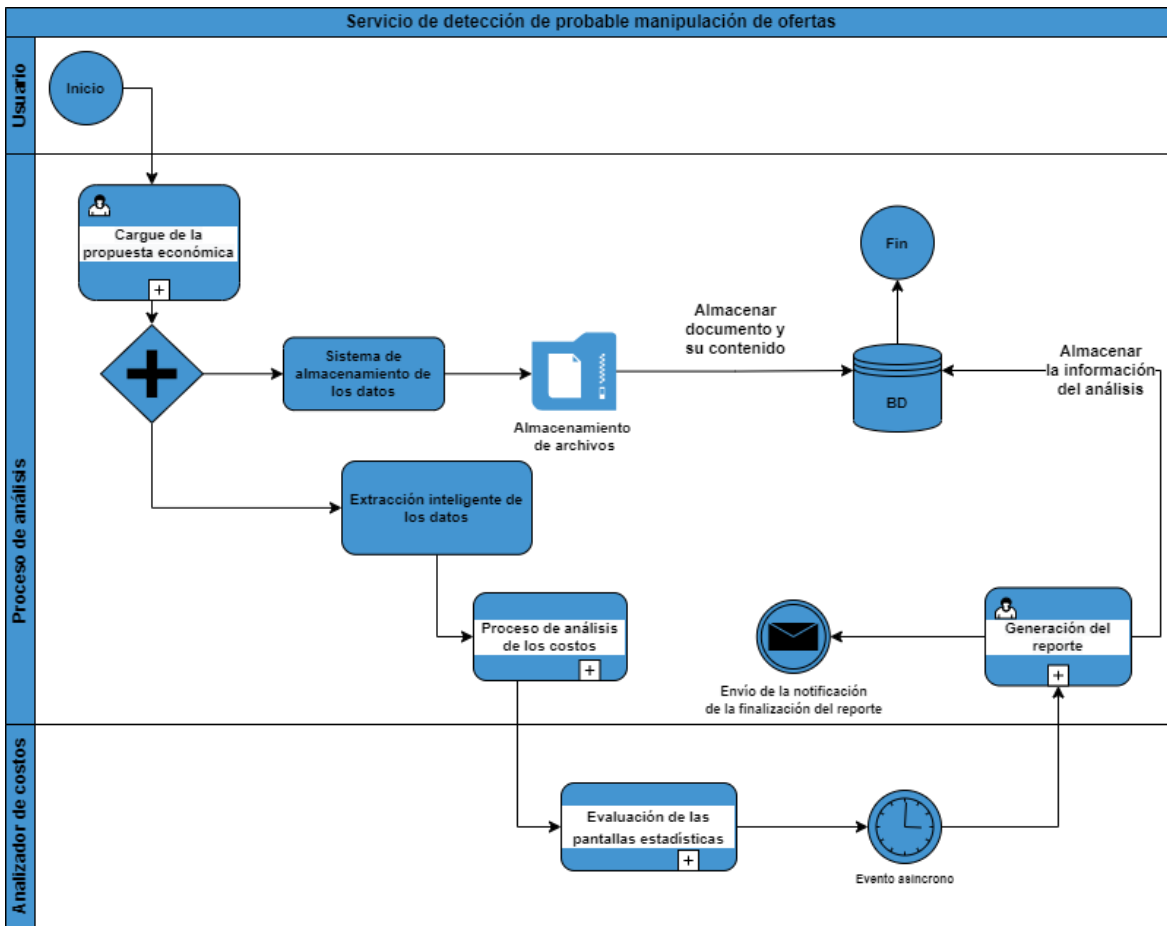


Figura 5. P.N. del servicio de detección de probable manipulación de ofertas

### 3.2.3.1. Extracción inteligente de datos

Como se mencionó anteriormente, para la creación de este modelo, y en especial, para la implementación de este servicio, se parte de la necesidad de tener una información estructurada, y para este caso en específico lo planteado en la **Tabla**

7. Sin embargo, y partiendo de la necesidad de poder almacenar información de anteriores procesos de licitación, y con ello realizar análisis post mortem que permitan definir nuevas estrategias para abordar este delito o nuevos tipos de delitos, se hace necesario contar con un método de extracción inteligente de datos sobre los documentos, dado que estos datos por lo general no están estructurados ni estandarizados como se pretende que estén al implementar el modelo propuesto en esta investigación.

Par lograr esto, en la literatura se pueden encontrar algunos mecanismos para llevar a cabo esta tarea, como algoritmos automáticos de extracción de datos [52] encargados específicamente de la extracción de datos encapsulados en tablas, lo cual se acopla a la necesidad de este servicio, sin embargo, dado el caso que los datos de los costos asociados a un producto/servicio no se encuentren formateados dentro de una tabla, algunas alternativas a contemplar para llevar a cabo la extracción y siguiendo con lo usado en la literatura, es el uso de técnicas de aprendizaje automático supervisado para la extracción de información de documentos que no hacen uso de un análisis lingüístico sofisticado, basados en aprendizaje de reglas y en algoritmos de clasificación [53].

Aparte de las propuestas académicas e investigativas, en el marco del software libre se pueden encontrar aplicaciones como ANTLR (<https://www.antlr.org/>) que a partir de filtros o reglas permiten identificar nombres propios (ciudades, personas, empresas, entre otros), valores y demás que son requeridos por este servicio y que muy seguramente deben ser usados en conjunto con otras técnicas inteligentes. En el caso de que los datos se suministran con el formato establecido en archivos de Excel o similar, el proceso se hace de una forma muy sencilla y verificando principalmente que las cantidades correspondan con lo solicitado y los valores (precios) no tengan problemas de cambio de formato o regionalización (separador de decimales basado en puntos cuando se tiene configurada la coma, por ejemplo).

Para este caso particular, los datos a extraer en este paso son los relacionados con un producto/servicio en particular que se está ofertando, entiéndase entonces como los valores numéricos de costos, los datos alfanuméricos del identificador y los datos alfanuméricos del detalle del producto/servicio, como las especificaciones técnicas del mismo (marca, modelo y detalles que puedan ser útiles para llevar a cabo la comparación con otras propuestas). Sin importar la técnica que se use y la alta tasa de éxito que se pueda llegar en el proceso automático de extracción, se debe evaluar la necesidad de un paso de revisión por parte de un humano para garantizar la total calidad de los datos extraídos.

### **3.2.3.2. Sub-Proceso de análisis de los precios (Evaluación de las pantallas estadísticas)**

Este proceso es el paso fundamental del servicio, en éste se realiza el análisis y se determina la probabilidad de presentarse el delito de manipulación de ofertas dentro de una licitación, para llevar a cabo esta tarea, y siguiendo con lo trabajado en la literatura, en este problema, se puede hacer uso de la aplicación de pantallas

(índices estadísticos específicos), con el propósito de agrupar aquellas propuestas que se asemejan en términos de costos.

En línea con lo anterior, a continuación, se exponen algunas pantallas utilizadas en la literatura como indicadores que sirven de prueba para señalar la probable presencia del delito de manipulación de ofertas en proceso de contratación. El conocimiento de estos indicadores, junto con su correcta aplicación puede desincentivar la existencia de acuerdos colusorios de manipulación de ofertas, así como de presencia de carteles [54], delito que se trata más adelante en la **sección 3.3** de este documento.

**Indicadores de varianza:** Estos indicadores sirven para analizar el comportamiento de la varianza en las ofertas mediante el uso de estadísticas simples como la desviación estándar o el coeficiente de variación. Este tipo de indicadores basan su funcionamiento en la predicción, contemplando la presencia de una disminución en la varianza de las ofertas económicas presentadas en un proceso de contratación cuando estas se relacionan con la presencia de carteles dentro del proceso.

**Coeficiente de variación:** Este coeficiente se define como la razón entre la desviación estándar y la media, es muy usado debido a su característica de invariabilidad de escala (se puede utilizar para comparar procesos de licitación de diversos valores). Según [10] obtener valores más bajos para este indicador señalaría una conspiración de manipulación de ofertas, en comparación con un proceso competitivo. El coeficiente de variación se calcula una vez para cada licitación ( $t$ ) donde la desviación estándar ( $\sigma_t$ ) de los valores de las ofertas presentadas por los diferentes ofertantes es dividida sobre el cálculo de la media aritmética ( $\mu_t$ ) de estas, tal y como se muestra en la **Ecuación ( 1 )**.

$$\text{Coeficiente de variación } (CV_t) = \frac{\sigma_t}{\mu_t} \quad ( 1 )$$

En el coeficiente de variación, la media ( $\mu_t$ ) tiende a aumentar cuando existe manipulación de ofertas dado que algunos de los proponentes presentarán ofertas más altas con el objetivo de aumentar sus ganancias. Por otro lado, la desviación estándar ( $\sigma_t$ ) tiende a disminuir dado el aumento entre el límite inferior y la estabilidad de la oferta del límite superior.

**Curtosis o apuntamiento:** La curtosis se puede entender como una medida estadística que determina el grado de concentración que presentan un conjunto de valores alrededor de la zona central. Esta medida permite conocer la homogeneidad de los datos de una variable. La curtosis ha sido utilizada para la detección del delito de manipulación de ofertas cuando las licitaciones presentan un escalado inteligente con un factor común a los costos reales, es decir sin grandes variaciones entre la oferta más alta y la más baja, lo cual complica en gran medida la detección del delito con el indicador de coeficiente de variación, dado que este permanece sin cambios significativos.

La **Ecuación ( 2 )** muestra el cálculo de la curtosis para una licitación  $t$ , donde  $n$  denota el número de ofertas totales de la licitación,  $b_{it}$  se refiere a la oferta  $i$  de la licitación,  $\sigma_t$  la desviación estándar de las ofertas en la licitación, y  $\mu_t$  la media aritmética de las ofertas de la mencionada licitación.

$$Kurt(b_t) = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \left( \frac{b_{it} - \mu_t}{\sigma_t} \right)^4 - \frac{3(n-1)^3}{(n-2)(n-3)} \quad (2)$$

Un elevado valor para este indicador señalaría una alta probabilidad de presencia de manipulación de ofertas en el proceso de licitación, las ofertas que contribuirán a aumentar este indicador serán aquellos que se encuentren ubicados a más de una desviación estándar de los datos de la muestra [54].

**Indicadores de cobertura:** Estos indicadores son muy utilizados en la literatura debido a su fácil implementación y obtención de buenos resultados, su propósito es analizar el comportamiento de la diferencia entre la oferta ganadora y las demás participantes de una licitación. Si se asume que en un proceso de licitación todos los proponentes acordaron realizar una manipulación de ofertas, los perdedores del proceso se aseguran de realizar ofertas excesivamente altas para garantizar su exclusión del proceso y así garantizar la escogencia del licitante previamente elegido por el grupo. Esto implicaría que la distancia entre las ofertas perdedoras será significativamente inferior en comparación con la oferta ganadora [54].

**Diferencia porcentual entre las dos ofertas más bajas de una licitación:** Este indicador consiste en obtener un valor porcentual equivalente a la diferencia entre la oferta ganadora y aquella que quedó en segundo lugar en un proceso de licitación. Como se mencionó en la descripción general de los indicadores de cobertura, es coherente afirmar que los proponentes de una licitación manipulan los precios con el objetivo de asegurar la elección del ganador, incrementando la distancia existente entre dicha oferta con las demás participantes. Como resultado, en caso de manipulación de ofertas, la diferencia porcentual tiende a aumentar entre las dos ofertas más bajas de una licitación. Según [10], luego de una recopilación de informes de testigos en casos de manipulación de ofertas regularmente se fija una distancia de cobertura del 3 al 5% entre la primera y la segunda oferta más baja presentada por el cártel. La **Ecuación ( 3 )** muestra el cálculo de la diferencia porcentual entre las dos ofertas más bajas en un proceso de licitación  $t$ .

$$Diferencia\ porcentual = \frac{b_{2t} - b_{1t}}{b_{1t}} \quad (3)$$

**Diferencia relativa:** Existen diferentes razones para explicar por qué las diferencias en las ofertas perdedoras tienden a disminuir. En primer lugar, los proponentes no quieren elevar los costos de su oferta demasiado, ya que una empresa que presenta ofertas demasiado altas puede ser una alerta para la entidad solicitante de los bienes o servicios. En segundo lugar, las ofertas dentro una manipulación tienden a replicar un proceso de competencia real, por lo que las ofertas perdedoras difieren



poco de la ganadora. Dado lo anterior y contemplando que, las ofertas de cobertura (ofertas creadas para perder y asegurar una ganadora previamente seleccionada) se basan en la oferta del ganador y pueden estar próximas entre sí. Esto sugiere que las diferencias entre las ofertantes perdedoras (ofertas de cobertura) disminuyen entre sí. Contemplando el indicador anterior, lo previamente mencionado y teniendo en cuenta que la diferencia entre las dos ofertas más bajas es importante y la diferencia entre las ofertas de cobertura es pequeña se construye el indicador de diferencia relativa [54] que se presenta en la **Ecuación ( 4 )**.

$$Diferencia\ relativa = \frac{b_{2t} - b_{1t}}{\sigma_{t,ofertas\ perdedoras}} \quad ( 4 )$$

Frente a la presencia de manipulación de ofertas se esperaría que la diferencia entre las dos ofertas más bajas sea alta, en comparación con un proceso de licitación competitivo, así mismo, la desviación estándar entre las ofertas perdedoras tiende a no ser muy elevada, dado que estas ofertas de cobertura tienden a pactar un precio alrededor de un valor estándar. Así pues, el valor de la diferencia relativa es un indicador porcentual de la probabilidad de la presencia del delito de manipulación de ofertas, si este valor tiende a uno indica una mayor probabilidad de presencia del delito, por otro lado, para procesos de licitación competitivos se esperaría que este valor tienda a cero.

Además, en la literatura recomiendan utilizar una mejora al anterior indicador, aplicando una normalización de este, en donde la diferencia de las dos ofertas más bajas es dividida por el promedio de las diferencias entre las ofertas adyacentes, tal y como se muestra en la **Ecuación ( 5 )**

$$Diferencia\ relativa\ Normalizada = \frac{b_{2t} - b_{1t}}{\frac{(\sum_{i=1}^{n-1} b_{jt} - b_{it})}{n - 1}} \quad ( 5 )$$

Donde  $b_{it}$ ,  $b_{jt}$  son ofertas adyacentes. Al contar con los valores de las ofertas ordenados de forma creciente. Un valor mayor que 1 indica que la diferencia entre la segunda y la primera oferta más baja es mayor que la diferencia promedio entre todas las ofertas adyacentes en una licitación.

Los indicadores de cobertura están destinados para aquellas licitaciones en donde todos los proponentes participan de la manipulación de las ofertas, una combinación de indicadores de cobertura junto con indicadores de varianza son una buena estrategia para identificar la presencia de este delito en aquellos procesos en donde no todos los proponentes confabularon para la fijación de los precios de las ofertas en un proceso de licitación. La presencia de este delito de forma parcial puede conllevar, entonces a la presencia de un grupo de ofertantes debidamente organizados para la consecución de licitaciones, lo que conllevaría a la presencia de un cartel de licitación, delito que es abordado más adelante en este documento (ver la **sección 3.3**).

### 3.2.4 Salida

Similar al servicio de detección de similitudes, el objetivo de este servicio es proporcionar al usuario final del modelo, un índice de probabilidad, basado en los indicadores presentados en la sección anterior, que le indique la probable existencia del delito de manipulación de ofertas, pero esta vez no por cada propuesta económica presentada sino de forma global y relacionado con el proceso de licitación. Es preciso comentar que además de obtener este índice global se debe calcular la probabilidad de existencia de este delito para cada una de las propuestas con el objetivo de buscar identificar las propuestas que están implicadas o no en el probable delito. Para suplir esta necesidad se propone la implementación del algoritmo de procesamiento del lenguaje natural **LexRank** (basado en PageRank) [55] el cual ha sido ampliamente usado para investigaciones relacionadas con redes sociales y generación automática de resúmenes. Este algoritmo trata los objetos a estudiar como nodos y los enlaces representan las relaciones entre ellos. Los autores del algoritmo plantean la hipótesis de que, si un nodo es muy similar a muchas de los otros, este se puede considerar como el más central.

En un proceso de contratación se esperaría que las ofertas económicas presentadas por los ofertantes sean, en un grado considerable similares o centrales. Las ofertas que se alejen, considerablemente de este centro pueden, en algunos casos y soportado con la información de las pantallas estadísticas mencionadas previamente, generar alerta de probable participación en el delito de manipulación de ofertas. Con esta hipótesis, en esta investigación se plantea el uso de este algoritmo para otorgarle un índice a cada oferta económica con respecto a su probabilidad de pertenencia en el delito de manipulación de ofertas. El algoritmo se muestra en la **Figura 6**. La primera tarea realizada por este (pasos 1 al 11) es la generación de la matriz de similitud de cosenos para cada una de las ofertas, seguidamente se aplica un umbral con el propósito de eliminar aquellas relaciones débiles entre las ofertas (nodos del grafo), es decir, aquellos vértices con similitud de coseno que no supera un valor dado, así mismo el algoritmo calcula el grado de centralidad de cada nodo contando el número de propuestas que tienen relación de similitud con otra propuesta después de haber sido aplicado el umbral. Se debe tener en cuenta que un valor de umbral demasiado bajo puede tener en cuenta en forma errónea similitudes débiles, mientras que un valor demasiado grande puede eliminar muchas de las relaciones de similitud en el grupo de ofertas.

Luego de esto se realiza una normalización (con ello cada fila se puede ver como un vector de cambios de estado con probabilidades que suman 1) de cada fila de la matriz de similitud. Esto se logra dividiendo cada elemento de la matriz por el grado de la fila en la que se encuentra (pasos 12 al 16). Seguido se continúa con la tarea de garantizar que la matriz estocástica sea irreducible y aperiódica, aplicando un factor de amortiguamiento a cada celda de la matriz (pasos 17 al 21).

Finalmente se aplica la función **PowerMethod** usando una tolerancia de error específico, por ejemplo 0.0001. El método hace uso de la propiedad de convergencia de las cadenas de **Markov** para calcular la distribución estacionaria

de un arreglo. En cada iteración, el vector propio se actualiza multiplicándolo con la transpuesta de la matriz estocástica y gracias a que el arreglo fijado anteriormente es irreducible y aperiódico se garantiza que este método no se queda estancado en un bucle. Con ello al final el algoritmo retornará un vector con un valor de centralidad para cada oferta, valor que para esta investigación será utilizado como índice de probabilidad de participación en el delito de manipulación de ofertas.

---

**Entradas:** Arreglo de propuestas económicas  $s$  de tamaño  $n$ ; Umbral  $t$ ; Valor amortiguamiento  $dampingFactor$

**Salidas:** Arreglo  $L$  con los scores definidos por el algoritmo para cada propuesta económica

```
1: cosineMatriz[n][n] = calcular matriz de similitud cosenos
2: Para  $i = 1$  hasta  $n$  incremento en 1 hacer
3:     suma = 0
4:     Para  $j = 1$  hasta  $n$  incremento en 1 hacer
5:         Si cosineMatriz[i][j] >  $t$  entonces
6:             cosineMatriz[i][j] = 1
7:             suma++
8:         Si no cosineMatriz[i][j] = 0
9:         Fin si
10:    Fin para
11: Fin para
12: Para  $i = 1$  hasta  $n$  incremento en 1 hacer
13:     Para  $j = 1$  hasta  $n$  incremento en 1 hacer
14:         cosineMatriz[i][j] = cosineMatriz[i][j] / suma
15:     Fin para
16: Fin para
17: Para  $i = 1$  hasta  $n$  incremento en 1 hacer
18:     Para  $j = 1$  hasta  $n$  incremento en 1 hacer
19:         cosineMatriz[i][j] = ( $dampingFactor / n$ ) + (1 -  $dampingFactor$ ) *
20:     cosineMatriz[i][j]
21:     Fin para
22: Fin para
22:  $L$  = calcular el PowerMethod a cosineMatriz
23: Retornar  $L$ 
```

---

Figura 6. Pseudocódigo algoritmo LexRank adaptado de [55]

### 3.3 Servicio de detección del probable delito de organización de carteles y análisis del beneficiario final

La organización de carteles se desglosa a partir del delito de colusión y también se conoce como acuerdos explícitos entre competidores. Consiste básicamente en la realización de acuerdos o pactos entre empresas con el objetivo de 1) fijar precios, 2) repartir el mercado entre los miembros del cartel o 3) frenar la producción de un bien o servicio con el objetivo de suprimir competidores en un mercado. La presencia de carteles en un mercado conlleva a la existencia de monopolios permitiéndole a una empresa o grupo de empresas asociadas, el control absoluto de un mercado, afectando directamente al consumidor, cliente o comprador. Las grandes consecuencias que generan la organización de carteles en un mercado son

principalmente dos, y se relacionan entre sí. 1) los carteles eliminan la competencia leal en un mercado y 2) encarecen el bien o servicio ofrecido [56].

En un escenario ideal de mercado, se debe presentar una sana competencia entre las empresas que contienden entre sí, implementado para ello, esfuerzos y estrategias. Esta sana competencia conlleva una serie de beneficios, los cuales son extendidos en el sector público cuando el Estado acude al mercado para la adquisición de algún bien, producto o servicio. Por lo anterior las conductas que limiten la sana competencia en un proceso de contratación pública son sancionadas en Colombia conforme con lo dispuesto en la Ley 155 de 1959 y el Decreto 2153 de 1992 prohibiendo, entre otras cosas, los acuerdos que se presentan entre los proponentes para no competir.

En procesos de contratación pública, la presencia de carteles limita la rivalidad que debe existir entre los ofertantes de una licitación, causando así, que el proyecto a adjudicar se eleve en costos y se reduzca la posibilidad de adjudicación de este a ofertantes que no pertenecen al cartel, eliminando así la libre competencia del mercado, afectando la capacidad del estado para cumplir con sus fines. Así mismo, la presencia de carteles dentro de un proceso de contratación pública puede llevar a otras prácticas corruptas, como el pago de coimas a participantes del proceso.

En Colombia en particular, la Superintendencia de Industria y Comercio (SIC) se ha posicionado como la entidad que busca investigar y luchar contra la organización de carteles, logrando imponer multas por el orden de 23 millones de dólares, con una tendencia al alza de esas cifras según proyecciones de la OCDE. Según la SIC [57], los procesos de licitación son un ambiente propicio para la presencia de carteles. Problema que ha llevado a que diferentes entidades ofertantes se abstengan de participar de procesos de contratación con el Estado considerando que no hay una competencia justa, por ello, algunas entidades deben, y en algunos casos han venido, implementando mecanismos que garanticen la selección justa de ofertas. Así, se ha logrado identificar algunas formas en las que el delito de carteles se puede presentar en el proceso de contratación estatal, las cuales se mencionan a continuación:

- **Durante la elaboración de los pliegos de condiciones:** Está es una etapa vulnerable en los procesos de licitación pública, dentro de las prácticas más comunes referentes a la presencia de carteles se encuentra el intercambio de información entre funcionarios y entidades proponentes, acordando precios y condiciones con el objetivo de garantizar la adjudicación del proyecto.
- **Durante la presentación de ofertas:** Esta es la etapa principal donde se arman los carteles de contratación, y se relaciona estrechamente con el delito de manipulación de ofertas o colusión, enmarcando el delito en algunos esquemas de colusión como: 1) ofertas de resguardo, 2) supresión de ofertas 3) rotación de ofertas y 4) asignación de mercado (estos esquemas fueron explicados previamente en la **sección 3.2**).
- **Durante la ejecución del contrato:** En esta etapa, según experiencias recogidas por la SIC, señalan que, con el objetivo de ocultar la presencia de

carteles, acuden a estrategias, tales como la subcontratación, o la cesión de contratos, entre otras. Razón por la cual se argumenta la importancia de lograr identificar el beneficiario final de los proponentes presentes en una licitación.

Aunque la detección del delito de organización de carteles presenta algunas dificultades debido a la naturaleza clandestina con la que actúan sus participantes, algunas señales presentes durante el proceso pueden ayudar a su detección. A continuación, se listan algunos de las alertas que se pueden presentar en algunos escenarios de contratación pública [58].

#### **En la presentación de ofertas:**

- La mejor oferta es presentada comúnmente por un mismo ofertante.
- Algunos ofertantes solo ofertan en licitaciones en una cierta parte geográfica aun cuando cuentan con la capacidad para hacerlo en diferentes zonas.
- Empresas que, teniendo la capacidad de ofertar en una licitación de una entidad específica, se limitan a hacerlo, aunque si lo hacen para otras entidades.
- Se presenta la deserción inesperada de ofertantes en un proceso de licitación.
- Se logra identificar un claro patrón de rotación del proponente ganador en un conjunto de licitaciones.
- Existencia de proponentes recurrentes en procesos de licitación que no han ganado o rara vez son adjudicados.
- Presencia de consorcios o “uniones temporales” con entidades que cuentan, de por sí, con la capacidad de ofertar de forma individual.
- El proponente adjudicado desiste o cede el contrato y posteriormente resulta como subcontratista de este.

#### **En la documentación presentada:**

- Se presentan altos niveles de similitud entre las propuestas presentadas por un grupo de ofertantes.
- Semejanza en la forma de redacción de las propuestas presentadas por distintos ofertantes.
- Los fondos desde donde se realizan los pagos previos, como pólizas, provienen de una cuenta conjunta de un grupo de ofertantes.
- Existencia de coincidencias en la relación del personal que presentan distintos ofertantes.
- Ofertas de distintos proponentes presentan similares errores de cálculo, o valores coincidentes.
- Las presentaciones de las propuestas se hacen en un tiempo similar (al cierre de la licitación) por distintos ofertantes.

#### **Relacionadas con los precios:**

- Algunos proponentes presentan propuestas económicas muy diferentes en las licitaciones en las que participan, aun cuando el objeto y las características a licitar son semejantes.

- Margen de precio llamativo entre la propuesta vencedora y las demás.
- Empresas de la zona donde se está realizando la licitación presentan costos de transporte semejantes a empresas que no pertenecen a la zona.
- Existencia de variaciones mínimas entre las propuestas de los ofertantes.
- Se presenta una disminución significativa en los precios cuando un nuevo proponente entra en la licitación (probablemente no perteneciente al cartel).

Si bien, la presencia de alguna de las conductas listadas previamente no conlleve a la presencia de carteles dentro de un proceso de contratación. Probablemente la existencia de un patrón constante de conductas de un proponente o de varios dentro de una entidad o en diversas entidades estatales, puede ser un indicador más eficaz para la detección de este delito.

La detección de este delito es una problemática que se ha tratado a lo largo de los años, mayormente de una forma empírica y manual, sin embargo, dada la complejidad de esta amenaza, se hace necesario la implementación de estrategias automáticas que faciliten a las entidades gubernamentales la detección eficiente y temprana de esta actividad. Basado en esto, y a pesar de que según [54] (2018) hay pocas herramientas automáticas destinadas para ello, algunas investigaciones o proyectos se han dedicado a esta tarea. A continuación, se mencionan los trabajos previos más relevantes, ordenadas cronológicamente y agrupadas en: Métodos econométricos, métodos sociales, métodos estadísticos y computacionales.

### **Métodos econométricos:**

Algunos procedimientos manuales basados en test econométricos han sido desarrollados para llevar a cabo esta tarea. Una recopilación de los métodos más utilizados en esta área se presenta en 2002 [59] donde se hace una descripción detallada de los mismos destacando que el principal enfoque, a la hora de realizar una detección manual de este delito, es comprender primero el comportamiento competitivo y sus implicaciones. Esto a razón de que teniendo identificadas las características del comportamiento competitivo, es mucho más fácil detectar desviaciones de este tipo de comportamiento entre las empresas. En la investigación se recalcan las dos principales condiciones que se deben presentar para poder clasificar el proceso de licitación como competitivo. Estas condiciones son la independencia condicional y la intercambiabilidad entre empresas. Gracias a esto, los autores indican que, mediante algunas pruebas econométricas se puede probar estadísticamente si se cumplen estas condiciones. Si las pruebas indican que la licitación de un grupo de empresas viola estas condiciones, entonces hay una gran probabilidad de que se haya presentado un comportamiento de cartel.

Similarmente, en 2013 [9] se propone un procedimiento econométrico que puede usarse para detectar la presencia de carteles basado en ofertas pasivas (patrón de oferta que a menudo se utiliza para crear la ilusión de competencia). El método planteado fue aplicado a datos de licitaciones de construcción de carreteras en Rusia, mostrando que cinco proponentes presentaron un comportamiento de oferta pasiva. Finalmente, una de las ventajas de esta investigación es que, según los

autores, el método propuesto, permite la detección de carteles en procesos de licitación sin tener conocimiento previo de la estructura del mercado.

### **Métodos sociales:**

En 2004 [60] se analizó el impacto de un programa de clemencia sobre los incentivos dentro de los cárteles. La idea consistió en alentar a un miembro del cártel a confesar e implicar a sus coconspiradoras con pruebas contundentes sobre su acuerdo colusorio. Para ello se desarrolló un modelo simple de comportamiento de cártel en un proceso de licitación. Los resultados mostraron que el programa de clemencia logró prevenir la coordinación interna de los miembros del cártel.

En 2018 [7] se introdujo un modelo que utiliza teorías criminológicas para el análisis y prevención de carteles. El modelo está compuesto por mecanismos de gobierno, controles internos y análisis de señales de alerta. Finalmente, los autores recalcan que los funcionarios de contratación pública cuentan con la información necesaria para descubrir un comportamiento de carteles (aunque tienen que recopilar pruebas sobre una serie de ofertas), sin embargo, esta tarea es de alta complejidad por lo que se deben implementar estrategias automáticas para facilitar su acción.

### **Métodos estadísticos y computacionales:**

En 2012 [20] plantearon que una de las principales tareas a la hora de aplicar una estrategia automática de detección de carteles es extraer la información útil de las bases de datos disponibles, tarea que se puede volver compleja debido al gran volumen de información que se utiliza para correlacionar los datos. Para lograr esto, los investigadores proponen dos estrategias de minería de datos: agrupación y reglas de asociación, junto con un enfoque de múltiples agentes integrando estas estrategias en una herramienta software buscando atacar el problema de extracción de información. El enfoque permitió el descubrimiento de conocimiento explícito debido a que la herramienta desarrollada presentó muchas reglas de asociación que proporcionaron una identificación correcta del 90% de la formación de carteles, según la evaluación de expertos.

En 2018 [61] se presentó un enfoque para encontrar rastros de organización de carteles aplicando técnicas de minería de datos para la extracción de información de las bases de datos públicas de licitaciones del Gobierno Federal de Brasil. El enfoque propuesto utiliza reglas de asociación para encontrar relaciones entre los proponentes. El modelo se utilizó para desarrollar un prototipo que ha sido utilizado por los auditores del Ministerio público.

En 2019 [18] se combinan técnicas de aprendizaje automático con pantallas estadísticas calculadas a partir de la distribución de ofertas en licitaciones dentro del sector de la construcción en Suiza, con el objetivo de predecir la presencia de carteles de contratación a través de manipulación de ofertas. Los resultados sugieren que la combinación de aprendizaje automático y pantallas estadísticas es una herramienta poderosa para identificar este conjunto de delitos (carteles y manipulación de ofertas), ya que luego de aplicar el método de validación cruzada

como mecanismo de evaluación del rendimiento de este enfoque se encontró que el mismo logró clasificar correctamente más del 84% del total de los procesos de licitación como colusorios o no colusorios.

Un enfoque diferente para la detección de carteles fue desarrollado en 2019 [22]. Se propone un marco basado en análisis de redes (grafos) para identificar grupos de empresas que interactúan intensamente en un mercado e identificar su potencial colusorio según la topología de la red. Los autores crearon una topología de red de empresas en función de su comportamiento de co-licitación, basado en los datos de organización de las empresas, suministrados en el proceso de licitación. Con ello en la investigación se logró detectar grupos que interactúan entre sí midiendo, además, su cohesión y exclusividad, dos características, que pueden indicar un comportamiento de cartel. La propuesta fue probada en diferentes contratos adjudicados entre 2011 y 2016. Demostrando que los grupos que presentaron alta cohesión y exclusividad tenían una elevada probabilidad de presentar un comportamiento de cartel.

También en 2019 [62] se realizó una comparación de métodos de minería de texto y aprendizaje automático para desarrollar un modelo de detección de carteles en proceso de licitación pública en Brasil. El modelo implementado utiliza algoritmos de clasificación de texto muy comunes para fines de predicción: Bayes ingenuo (Naive Bayes), regresión logística y máquinas de soporte vectorial, siendo las dos últimas las que demostraron ser mejores para realizar predicciones relacionadas con la presencia de carteles.

En 2020 [23] se utilizó un enfoque diferente, en donde se establece la presencia de carteles en las licitaciones mediante el análisis de las direcciones IP de los miembros de una licitación pública. El documento establece una idea innovadora en el campo de la identificación de corrupción en el proceso de contratación pública, obteniendo resultados prometedores.

Recientemente, en 2021 [63] se propone un modelo basado en aprendizaje automático, utilizando pantallas estadísticas derivadas de la distribución de los precios ofertados por cada propuesta para detectar la presencia de carteles en los procesos de licitación. A manera de innovación los autores calculan pantallas algorítmicas basadas en la organización de coaliciones por parte de los ofertantes de una licitación. El método fue aplicado a tres conjuntos de datos diferentes de Japón, Suiza e Italia para los que se conocía la incidencia de carteles en contratación. En los tres países, se obtuvo en promedio unos buenos resultados de clasificación.

Por último y también en 2021 [64] luego de realizar una revisión de la literatura sobre el área, se propone el uso de una técnica de aprendizaje profundo que señala a los participantes de un cartel en un proceso de licitación, basado en sus interacciones de licitación por pares con otras empresas. Los autores, combinan redes neuronales convolucionales (CNN) para el reconocimiento de imágenes con gráficos, de las interacciones de ofertas por pares de empresas en varias licitaciones. Basado en



los datos de adquisiciones japoneses y suizos, construyeron gráficos para licitaciones que presentaban organización de carteles y para las que no. Un subconjunto de gráficos fue utilizado para entrenar la red neuronal de modo que aprenda a distinguir entre patrones de licitación competitiva y de presencia de carteles. Los resultados arrojados por el modelo obtuvieron una precisión media de alrededor del 90%. Finalmente, los autores recalcan el hecho de que, gracias a los resultados obtenidos por el modelo y teniendo en cuenta el éxito que ha tenido el uso de técnicas de aprendizaje profundo en muchos dominios, el enfoque propuesto por la investigación tiene el potencial de superar a otros enfoques cuantitativos para señalar la presencia de carteles y es una de las aplicaciones más novedosas de las CNN en economía hasta la fecha.

En resumen, se logra identificar como, similar a las anteriores prácticas indebidas tratadas en este documento, la identificación de la presencia de carteles ha sido una tarea que se ha venido trabajando a lo largo de los años. Diversas investigaciones han tratado de abordar el problema con diferentes metodologías, las cuales han ido evolucionando a lo largo de los años, pasando de test econométricos a técnicas estadísticas y últimamente, mezclando estas últimas con técnicas computacionales, siendo el uso de técnicas de aprendizaje profundo el auge de esta problemática, aprovechando los amplios beneficios que se pueden obtener de esto.

El futuro de las soluciones a esta problemática, consisten en el aprovechamiento que se le pueda dar a los grandes volúmenes de datos y documentos referente de los proponentes y las ofertas que se puedan ir almacenando por cada licitación realizada, esto similar a los servicios ya tratados en este documento, necesita un modelo previamente definido, en donde, primero se garantice el almacenamiento sistemático de dicha información. Luego, se hace necesaria la aplicación de técnicas de minería de datos, inteligencia artificial y demás herramientas de TI que permitan sacarle provecho a los datos con el objetivo de atacar la problemática de la identificación del delito de carteles en los procesos de contratación pública.

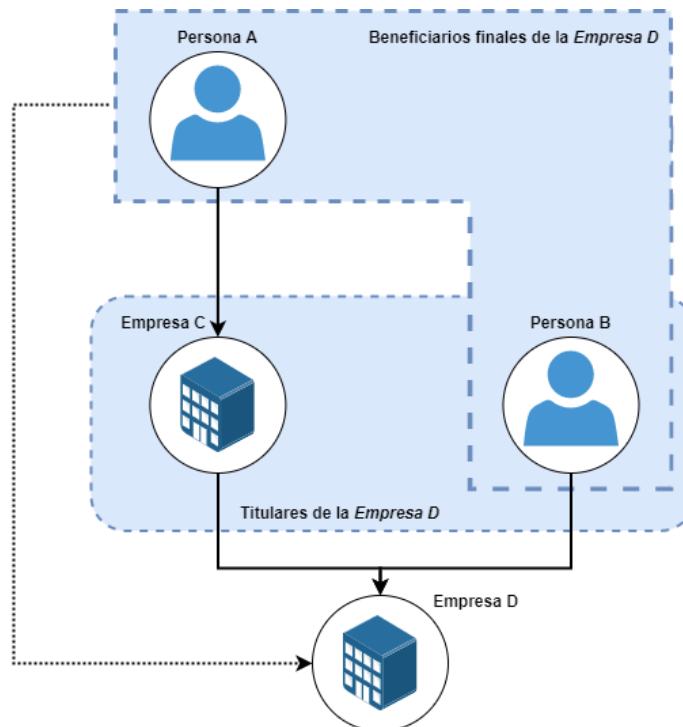
Por otro lado, la correcta implementación de una técnica que permita la identificación de la organización de carteles dentro de un proceso de licitación mediante la información detallada de los proponentes, podría ayudar, además, con la identificación del Beneficiario Final (BF) de cada participante, entendiéndose como Beneficiario Final a la persona natural propietaria y quien tiene el control de la empresa y/o persona que está participando en una transacción, en este caso particular, el proponente de una convocatoria de contratación.

La **Figura 7** es una representación gráfica que expresa cómo funciona la titularidad de una empresa. En esta figura, la Persona B y la Empresa C son los titulares legales de la Empresa D. La Persona A es la titular legal de la Empresa C. La Persona A y la Persona B son las beneficiarias finales de la Empresa D. La Persona B ejerce su titularidad de forma directa, mientras que la Persona A lo hace de forma indirecta a través de la Empresa C. La Empresa C no puede ser una beneficiaria final porque no es una persona física sino jurídica.

La anterior definición expresa que no necesariamente el BF está atado a una conducta ilícita, sin embargo, en diversas oportunidades las personas naturales han empleado una o varias compañías para esconder su identidad y origen de los recursos, bien sea para dar apariencia de legalidad de recursos de origen ilícito o para desarrollar una conducta ilícita [65]. Es por ello, que una correcta identificación del BF puede ser clave para evitar probables actos corruptos dentro de un proceso de contratación, prescindiendo así, negocios a ciegas, con el objetivo de evitar vínculos con empresas fachadas, testaferros, y/o con dueños vinculados previamente con actos de corrupción.

La correcta identificación de los BF puede ayudar a prevenir la corrupción en un proceso de contratación en dos modos fundamentales, 1) Detectando conflictos de interés no divulgados con anterioridad y 2) Generando alertas sobre probables carteles y manipulación de ofertas. Para esto, cada vez son más los países que han comenzado a implementar prácticas de transparencia en materia de beneficiarios finales para los procesos de contratación pública, siendo Colombia uno de ellos [4].

Para ello, la principal recomendación realizada por el Grupo de Acción Financiera Internacional (GAFI), las entidades deben tomar medidas para obtener la información del BF, lo cual se puede lograr mediante el diligenciamiento de formularios o la recopilación de información mediante la realización de entrevistas.



**Figura 7.** Beneficiarios finales

A continuación, se presentan algunas señales de alerta recopiladas por ASOBANCARIA [66] que podrían ser útiles para la correcta identificación del BF:

- Cambios en los nombres de las personas jurídicas o sociedades sin que exista aparente necesidad de hacerlo.
- Sociedades constituidas en paraísos fiscales que no exigen el registro que identifique los accionistas de una persona jurídica.
- Desconocimiento por parte del representante legal o administrador sobre quiénes son los accionistas de la respectiva sociedad.

Según GAFI [67] conocer el BF en una transacción en una obligación de medio y no de resultado, por lo que las entidades deben implementar cualquier tipo de estrategia que este a su alcance para cumplir con este objetivo llevando a cabo cualquier medida razonable como implementación de metodologías, algoritmos, etc.

Definir e identificar correctamente la topología que se presenta en una convocatoria de contratación a razón de sus proponentes es un tema clave, tanto para identificar la probabilidad de presencia de uno o más carteles, como para ayudar a la correcta identificación del beneficiario final de cada uno de los proponentes, pero para ello, es clave contar con un almacenamiento sistemático de la información recopilada de los datos generales y personales de los proponentes.

### 3.3.1 Datos de entrada requeridos

En un proceso de contratación pública, un proponente primero estudia detalladamente el pliego de condiciones verifica que cumple con los requisitos planteados para participar en la licitación. Cumpliendo los requisitos y contando con la capacidad para cumplir con el pliego de condiciones, procede a elaborar la propuesta y la entrega en los plazos establecidos. Dicha propuesta debe cumplir con claridad las exigencias realizadas por la entidad pública, así como cumplir con todos los formatos y anexos, el no cumplimiento de estos llevará a la exclusión del proponente del proceso de licitación.

Basado en lo anterior, y sabiendo que, por lo general en un proceso de licitación, la primera etapa de documentación hace referencia a los documentos del proponente como tal, los cuales están destinados para que la entidad verifique la capacidad de contratar del proponente contemplando aspectos jurídicos, financieros, económicos y técnicos. Para la implementación de este servicio, se parte de los documentos que se deben presentar para poder llevar a cabo una identificación de la probabilidad de existencia de carteles dentro del proceso de contratación, así como de la identificación del beneficiario final de una entidad ofertante.

**Registro único de proponentes (RUP):** Este documento, enmarcado en la Ley 80 de 1993 y establecido en la Ley 1150 de 2007 como principio del régimen de contratación estatal, es de carácter obligatorio en los procesos de licitaciones de contratación pública y es administrado por las cámaras de comercio regionales, la información registrada en el RUP debe estar debidamente soportada con sus respectivos anexos. El RUP habilita a personas naturales o jurídicas para la celebración de contratos con entidades estatales colombianas. Este es un registro que suministra, además de los datos personales, la información de un proponente

relacionado con su capacidad jurídica, experiencia, capacidad financiera y organizacional. En él se registra información relacionada con los servicios, bienes u obras de contratos (incluyendo las multas y sanciones) en los cuales participa el proponente, siendo esta información actualizada periódicamente. En el RUP entonces se puede encontrar la siguiente información:

- **Información general:** En este apartado se registra información relacionado con el nombre o razón social del proponente, la identificación, indicación del tamaño de la organización (micro, pequeña, mediana, grande), domicilio principal, la indicación de los bienes, obras y servicios que ofrecerá a las entidades estatales, basado en el clasificador de bienes y servicios. Además, para personas jurídicas se cuenta con información referente a la formación de la personería jurídica, así como los datos del representante legal.
- **Experiencia:** En este apartado se registra la información relevante sobre los contratos (en ejecución y ejecutados) que hayan sido celebrados por el proponente con alguna entidad estatal. En caso de nuevos proponentes, en este apartado se puede registrar experiencia de los accionistas, socios, consorcios o uniones temporales en donde haya tenido participación. Este registro es clave para determinar la idoneidad del proponente con respecto a una oferta realizada por una entidad Estatal.
- **Capacidad jurídica:** Las personas jurídicas deben informar cuál es su objeto social principal, así como la capacidad jurídica que tienen para celebrar contratos sobre los bienes, obras y servicios requeridos por el Estado.
- **Capacidad financiera:** En este apartado se registran los indicadores que permitan medir la fortaleza financiera del proponente basado en los últimos tres (3) años. Estos indicadores son:
  - Índice de liquidez: activo corriente/pasivo corriente
  - Índice de endeudamiento: pasivo total/activo total
  - Razón de cobertura de intereses: utilidad operacional/gastos de intereses
- **Capacidad organizacional:** En este apartado se registra la información que permitan medir el rendimiento de las inversiones del proponente en los últimos tres (3) años. Los indicadores medibles son:
  - Rentabilidad del patrimonio: utilidad o pérdida operacional/patrimonio
  - Rentabilidad del activo: utilidad o pérdida operacional/activo total

Finalmente, es clave mencionar que, por Ley, en Colombia cualquier persona puede consultar de forma libre la información contenida en el RUP. Estos datos, además de ser sumamente útiles para detectar nexos o relaciones entre los proponentes también son una herramienta para identificar el historial de proyectos trabajados por cada proponente, lo que es de suma importancia para poder seleccionar el candidato idóneo en un proceso de contratación.

Teniendo en cuenta lo anterior, para la implementación de este servicio, se parte entonces de la existencia del registro RUP, siendo este, junto con sus anexos y el servicio de indexación, detallado previamente en esta investigación, las fuentes de información que se utilizarán para llevar a cabo la tarea de detección de carteles, identificación del beneficiario final e idoneidad del proponente.

Buscando suplir la falta de datos sobre la conformación de las empresas, se plantea el uso de herramientas en línea encargadas de recopilar datos sobre empresas alrededor del mundo, proporcionando identificadores únicos que se pueden usar para vincular información sobre corporaciones, desde informes financieros hasta informes de licencias y contaminación. Una de estas herramientas, la cual ha sido ampliamente utilizada para este fin es el software **OpenCorporates** (<https://opencorporates.com/>) que además de ser de uso libre, cuenta con información de aproximadamente 40 millones de empresas en todo el mundo.

### 3.3.2 Procedimiento

El servicio de detección de carteles propuesto en esta investigación (que además busca brindar información relevante sobre el beneficiario final y la idoneidad de los proponentes), consiste en la generación de una topología de red, haciendo uso de la información relevante de los proponentes, junto con el histórico de participaciones en procesos de contratación, y el posterior análisis de dichas topologías para la detección de patrones de comportamiento que permitan alertar la formación de carteles dentro de un proceso de licitación. Además, gracias a la creación de la red, se puede, en algunos casos, lograr detectar el beneficiario final de un proponente y encontrar relaciones con los demás beneficiarios finales.

A diferencia de los anteriores servicios expuestos en este documento, para el planteamiento de este servicio, se debe contar, primero con datos históricos de contratación, así como la información histórica de licitaciones adjudicados para cada proponente (información que se registra en el RUP), y que, como se menciona previamente en esta investigación, la información de los procesos de contratación se vaya almacenando en la herramienta que implementará los servicios. Con ello, se puede construir de forma más eficiente la topología de red de los proponentes presentes en una convocatoria de contratación.

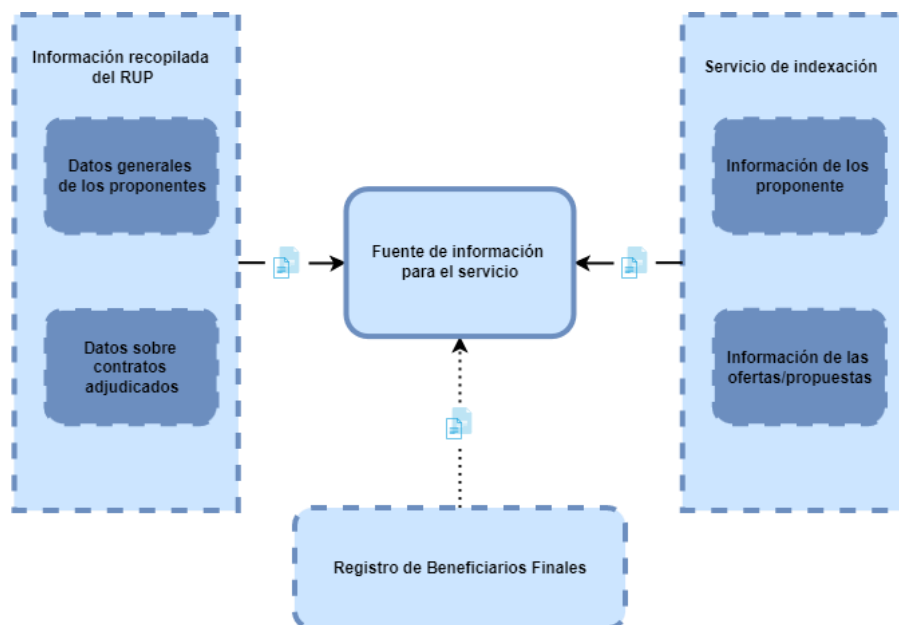
La **Figura 9** presenta un diagrama de procesos de negocios que modela el funcionamiento del servicio, en ella se logra apreciar que el proceso inicia con una etapa de recuperación de la información necesaria para el posterior análisis (en este servicio se parte de la idea que tanto la información de los proponentes, como de los documentos solicitados para el proceso de licitación han sido cargados).

El proceso de recopilación de datos se muestra en la **Figura 8**, el cual consiste en extraer la información general de las organizaciones presentes en un proceso de contratación, tanto del RUP como del servicio de indexación, explicado previamente en este documento. Esta información incluye nombre o razón social, identificación, representante legal, etc. Además, el proceso también extrae información de los contratos adjudicados a cada proponente (fuente RUP) y de las propuestas ejecutadas por este, almacenadas en el servicio de indexación. Por otro lado, se plantea, como adicional, la recopilación de información de los beneficiarios finales de las fuentes de información estatal (Proyecto de Ley que en Colombia se encuentra en proceso).

El proceso consiste en la creación de gráficos que representan la interacción presentada en un proceso de contratación entre un proponente y los demás basados en sus ofertas económicas, siguiendo las ideas presentadas en [19, 20], usando para ello una transformación mínima-máxima (normalización de datos) para cada oferta, tal y como se muestra en la **Ecuaación**  $b_{it}$  donde la oferta más baja y la más alta de un proceso de licitación  $t$  son representadas por  $b_{min,t}$  y  $b_{max,t}$  respectivamente. Con esto, cada oferta o propuesta económica toma valores entre 0 y 1 ( $b_{it} \in [0,1]$ ) siendo estos los valores que se dibujan en los gráficos mediante el cálculo de las coordenadas cartesianas resultantes de agrupar cada oferta (eje X) con las demás (eje Y).

$$Normalización (b_{it}) = \frac{b_{it} - b_{min,t}}{b_{max,t} - b_{min,t}} \quad (6)$$

Luego de ello, el modelo propone el uso de Redes Neuronales Convolucionales (CNN) debido a su alto desempeño en el análisis e imágenes, las cuales tienen como objetivo desarrollar modelos predictivos mediante la enseñanza de patrones sistemáticos de datos, para predecir la probable presencia de carteles en el proceso de contratación mediante el análisis de los gráficos generados previamente.



**Figura 8.** Proceso de recopilación de la fuente de información del servicio

Posterior al proceso de recopilación de la fuente de información, el modelo continúa con la etapa de identificación de carteles dentro del proceso de contratación, esta etapa se representa en la **Figura 9**. Este servicio usa lo propuesto en [64] por los buenos resultados obtenidos y la actualidad de la investigación realizada.

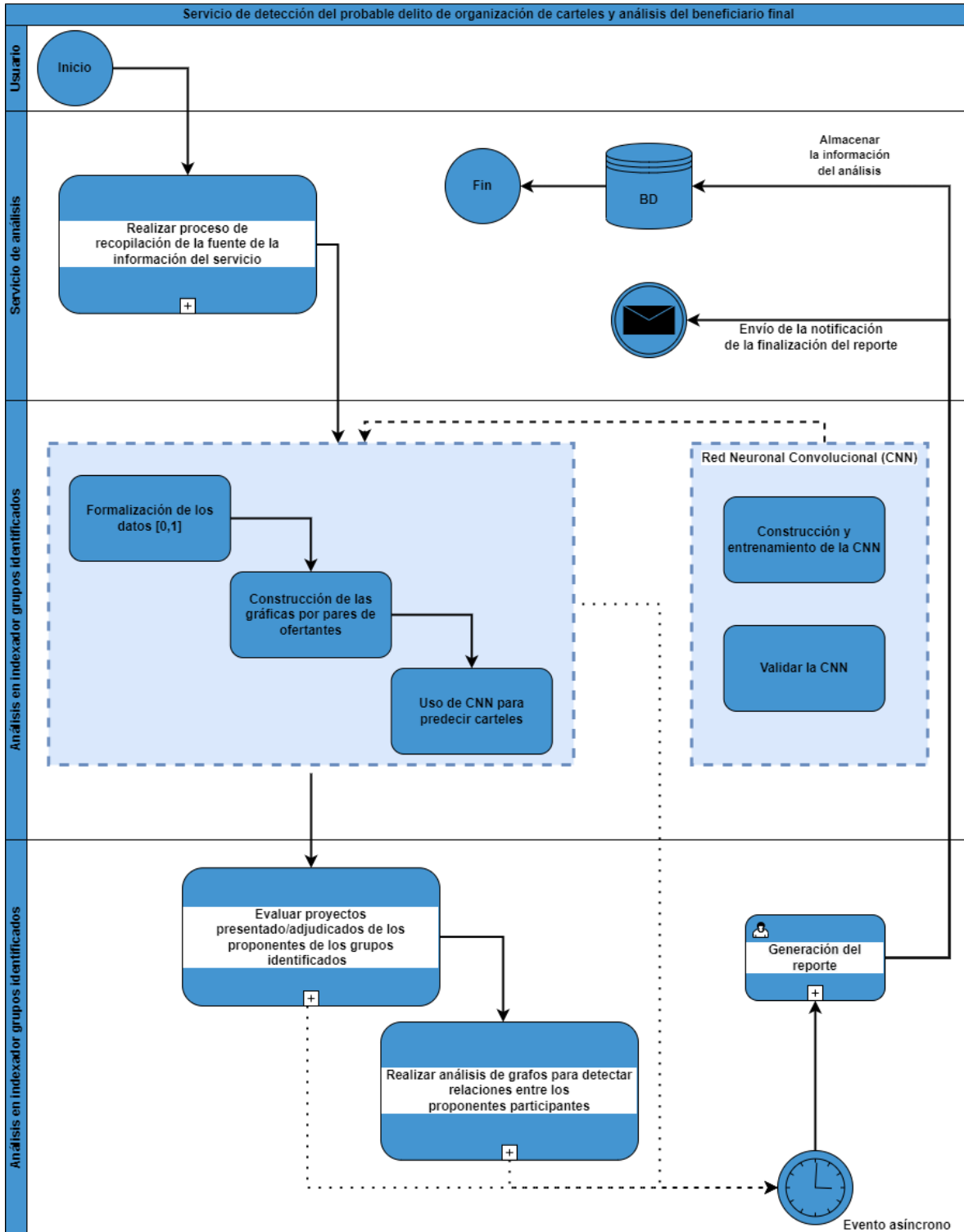
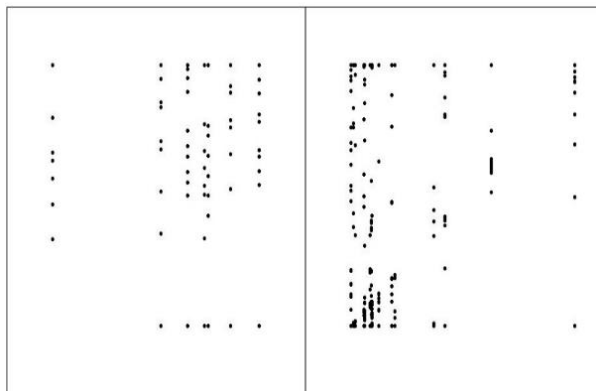


Figura 9. P.N. del servicio de detección de probable organización de carteles

La **Figura 10** muestra un ejemplo de dos imágenes, la de la izquierda representa un gráfico de un proceso de licitación con presencia de carteles y la de la derecha un proceso competitivo. Se logra apreciar como en la primera imagen la zona inferior

izquierda (la que debería contener las ofertas económicas más bajas) está vacía mientras se observa más interacciones por pares en la región superior derecha. Por otro lado, en la segunda imagen la agrupación de las iteraciones por pares de proponentes se acumula en la zona competitiva (inferior izquierda).



**Figura 10.** Gráfica de interacciones de proponentes. Tomada de [64]

En un siguiente paso, el modelo plantea una etapa de evaluación de los grupos identificados en el previo análisis de las gráficas generadas, esto haciendo uso de la información sobre el histórico de ofertas realizadas por los proponentes en diversas convocatorias almacenadas en el indexador. En esta etapa se busca entonces, identificar en que licitaciones previas los integrantes del grupo del cartel, previamente identificado, han operado, para con ello detectar el tipo de cartel que se ha organizado para lo cual se puede hacer uso de los indicadores abordados en la **sección 3.2.3.2** relacionada con el Sub-Proceso de análisis de los precios. Además, con la información recogida en el servicio de indexación y en el RUP se puede generar un reporte sobre la idoneidad de cada ofertante, basado en su experiencia previa en la presentación/adjudicación de proyectos.

Posteriormente, un análisis, mediante grafos de los proponentes participantes del cartel (identificado en el análisis de los gráficos) es propuesto con el objetivo de identificar las relaciones personales/laborales que existan entre ellos. Una profundización de este análisis puede, entonces, llevar a identificar los beneficiarios finales de cada empresa presente. Finalmente, una vez realizado el análisis, el proceso culmina con la entrega del reporte, dentro de un motor de base de datos y la generación de una notificación del estado final de la tarea al usuario.

### 3.3.3 Salida

Este servicio también brinda al usuario final del modelo un índice de probabilidad que indicaría la presencia o no de un cartel dentro del proceso de licitación, esta tarea es realizada gracias al análisis de las gráficas generado por la CNN, este análisis es generado a manera global sobre el proceso, sin embargo, gracias a la identificación de los grupos mediante las imágenes generadas se pueden identificar los proponentes que pertenecen a la organización de cartel, así, este servicio,



además de generar índices de corrupción, permite generar información relevante sobre cada uno de los proponentes presentes en el proceso de licitación.

Esta información, si bien no está en el alcance del modelo presentado puede ser tratada en servicios futuros que se puedan adicionar al modelo. La misma puede ser un análisis de idoneidad de cada proponente con respecto al tipo de propuesta ofertada, así como la identificación de las relaciones existentes entre cada uno de ellos y la identificación de sus beneficiarios finales.

Estos reportes son generados, tal y como lo muestra la **Figura 9** al final del proceso y los mismos pueden ser visualizados por parte del usuario en un reporte general.

### **3.4 Servicio de alerta ciudadana sobre irregularidades en procesos de convocatorias de contratación pública**

Como se ha mencionado anteriormente en esta investigación, la detección, identificación y documentación de corrupción, en especial en procesos de contratación pública, es una tarea difícil debido a la naturaleza clandestina de los diferentes delitos que se pueden presentar. Hasta el momento, se ha planteado como, desde una perspectiva tecnológica, es posible identificar y detectar algunos de los delitos más comunes en esta área, sin embargo, teniendo en cuenta que, según la OCDE [69] los procesos de contratación pública son un foco de presencia de diversos delitos de corrupción, se hace necesario contemplar una estrategia que permita identificar prácticas corruptas para las cuales no se tiene un servicio especializado.

La complejidad del proceso de contratación pública en las fases previas lo hace vulnerable a prácticas corruptas. Por ejemplo, durante las fases de licitación, es probable que participantes del mismo modifiquen las reglas de contratación para favorecer a los licitadores preferidos a cambio de sobornos [70]. Otro problema que se ha identificado en la etapa de definición de requisitos es la tendencia de los funcionarios públicos a diseñar la licitación adaptadas a un postor favorito, ya que estas partes formulan la licitación o convocatoria. Como resultado, el proceso de licitación se distorsiona en secreto. Esto da lugar a diversas prácticas corruptas discriminatorias que incluyen favoritismo, amiguismo, nepotismo y clientelismo, entre otras formas de corrupción [71].

En [72] mencionan que una de las prácticas corruptas que mejor suele ocultarse en procesos de contratación pública es el conflicto de interés, el cual es un hecho que debe ser reportado por los implicados, pero en algunos casos es omitido. Los autores mencionan que contar con una aplicación de denuncia ciudadana, que esté bien soportado, puede ayudar a evitar, desde el inicio de un proceso de contratación la presencia de posteriores delitos que afecten los dineros públicos. Generalmente la identificación de este tipo de delitos depende de denuncias, solicitudes de investigación, o profundas sospechas sobre un acto delictivo, y en la mayoría de los casos, estas tienen su efecto, mucho después de la adjudicación de los proyectos, conllevando así a pérdidas de dineros públicos que pueden ser bastante elevados.

En [30] mencionan que la detección temprana de los factores de riesgo y los factores causales que impulsan la incidencia de prácticas corruptas puede ayudar a mitigar la manifestación y los efectos asociados con la corrupción. Sin embargo, debido a la dificultad de medir la corrupción, se convierte en una tarea compleja el establecer medidas para detectar la probabilidad de vulnerabilidad de una licitación a la corrupción basado en diversos delitos, y más aún si estos son delitos que se ocultan en el comportamiento social de los participantes, para lo cual, el uso de denuncias públicas y una estrategia para, automatizar y sistematizar estas, puede ser una herramienta útil en esta tarea.

Según [73] para el desarrollo de índices de corrupción se hace necesario tener en consideración algunas variables de entrada que permitirán la creación de un modelo eficiente y riguroso. Algunas de estas variables han sido adoptadas en mayor grado en los servicios especializados formulados anteriormente. Entre dichas variables se destaca el constructo que hace referencia a la recopilación de datos de percepciones y experiencias (internas y/o externas) tanto del proceso de contratación, la entidad y los proponentes, para el cual se puede hacer uso de diversas estrategias para su recopilación, como lo son, encuestas de denuncia pública y/o anónima que, si bien, pueden ser categorizadas como imprecisas debido a que pueden ser marcadas como percepciones y, por lo tanto, pueden poseer un nivel apreciable de distorsión, estos datos, si vienen acompañados de un debido soporte [74], como en este caso de una serie de servicios especializados que se enfocan en la detección de delitos específicos, pueden ser representativos para alertar tempranamente de situaciones anómalas en procesos de contratación.

Según el informe de Transparencia Internacional [1] aproximadamente el 74% de los encuestados cree que la ciudadanía puede hacer una diferencia en la lucha contra la corrupción en Colombia, además de indicar que, aunque la ciudadanía tiene una alta disposición para denunciar casos de corrupción, los desmotiva los trámites que implica.

La falta de confianza hacia las entidades judiciales, las dificultades en los procesos de formalización de denuncia, el miedo a amenazas y presiones, son algunas de las principales razones por las cuales la población común ha disminuido su interés en presentar denuncias públicas sobre situaciones relacionadas con corrupción muy a pesar de que las instituciones han establecido mecanismos ágiles para recibir denuncias ciudadanas sobre hechos de corrupción dentro de sus instituciones (Ventanillas de recepción de denuncias, líneas telefónicas, aplicativos móviles, correos electrónicos, y sitios webs).

Con lo anterior se tiene una necesidad clara de brindar un servicio de denuncia ciudadana sobre los procesos de contratación vigentes, en el cual los ciudadanos puedan acceder o bien a nombre propio o de forma anónima y les permita registrar sus denuncias relacionadas sobre un proceso de contratación y/o los participantes involucrados en el, mediante la narración de hechos asociados a probables actos corruptos, esto mediante el suministro de sustentos probatorios.

Este servicio entonces tiene dos frentes, uno enfocado a la recopilación de las denuncias ciudadanas, el cual deberá ser accedido de forma pública y un segundo frente que permita procesar, relacionar y analizar las denuncias, el cual está enfocado a la entidad propietaria del modelo.

La recopilación de estos datos puede servir de: 1) apoyo a delitos atacados en los servicios previamente descritos, 2) identificación de nuevos delitos 3) herramienta de recopilación de información sobre los actores de los procesos de contratación y sus intereses, relaciones, entre otros.

### 3.4.1 Datos de entrada requeridos

Teniendo en cuenta los lineamientos del Departamento Nacional de Planeación de Colombia (DNP) que sugiere se tengan en cuenta los siguientes ítems para llevar a cabo la presentación de una denuncia:

- Presentar una relación clara, detallada y precisa de los hechos de los cuales se tiene conocimiento.
- Expresar cómo ocurrieron los hechos, el lugar y la fecha.
- De tener conocimiento, señalar quién o quiénes participaron en los hechos.
- Adjuntar las evidencias que sustentan el relato.
- En lo posible suministrar el nombre y dirección de residencia, teléfono y correo electrónico para un futuro contacto bien sea necesario para el avance de la denuncia o para mantener informado del curso de su denuncia.
- En el evento de preferir presentar la denuncia de forma anónima, asegurarse de que ésta amerite credibilidad acompañándola de evidencias que permitan orientar la investigación.
- Informar si los hechos han sido puestos en conocimiento de otra autoridad, indique cuál.

Con lo anterior el servicio de denuncia ciudadana de probables actos corruptos en procesos de contratación debe tener en cuenta los datos de entrada que se listan en la **Tabla 8**.

**Tabla 8.** Datos de entrada necesarios para el servicio de denuncia ciudadana

<b>Dato de entrada</b>	<b>Descripción</b>	<b>Obligatorio</b>
Tipo de solicitante	Indica el tipo de razón social de quien realiza la denuncia (Ciudadano, empresa)	Si
Datos personales	Información personal de quien denuncia, de ser ciudadano corresponde a nombres, apellidos, tipo y número de documento, teléfonos de contacto, dirección de domicilio y correo electrónico.  De ser Empresa corresponde a Razón social, tipo y número de documento, teléfonos de contacto, dirección de oficina y correo electrónico.	No

Autorización para respuesta de la denuncia	Campo que indica a qué medio se puede dar respuesta a la denuncia impartida, el mismo podría ser vía correo electrónico, teléfono o dirección física	No
Tipo de denuncia	Indica el tipo de denuncia que se desea radicar, bien sea enfocada al proceso de contratación, a un funcionario de la entidad o a un licitante del proceso	Si
Asunto de la denuncia	Similar a un asunto en un correo electrónico, este campo permite etiquetar la denuncia recibida	Si
Tipificación del delito a denunciar	Campo que indica el tipo de delito que se está denunciando, campo que es útil para la recopilación de denuncias similares	Si
Descripción de la denuncia	Campo en donde el denunciante plasma los hechos relacionados con presuntos actos de corrupción dentro del proceso de contratación	Si
Fecha de la denuncia	Fecha de cuando fue diligenciada la denuncia	Si
Anexos	Material probatorio que sustenta la denuncia	Si

Es importante mencionar que los campos marcados como no obligatorios tienen esta marca solo si se especifica que la denuncia es anónima.

### 3.4.2 Procedimiento

El gobierno electrónico (E-Government) es una tendencia que ha venido siendo implementado por instituciones públicas en los últimos años consiste en el uso de tecnologías de comunicaciones electrónicas, como Internet, para mejorar y promover el acceso de los ciudadanos a los servicios públicos [75]. Basado en esto, este servicio se propone como un modelo de gobierno electrónico basado en el enfoque de web 2.0 facilitando el intercambio de información, la interoperabilidad y la colaboración entre el usuario y la web. Permitiendo así recopilar y gestionar las denuncias ciudadanas sobre procesos de contratación estatal.

El modelo presentado en la **Figura 11** consta de cuatro etapas principales, las cuales son:

- **Recepción de la denuncia:** Proceso inicial del servicio, en donde se fija la forma de denuncia a realizar (personal o anónima) y se toman los datos de contacto, de ser el caso y los datos de la denuncia misma, junto con los anexos que la soportan, como enlaces relacionados, audios, imágenes, videos, entre otros. Al final se informa al denunciante del éxito o no de la recepción de la denuncia. La implementación de este servicio debe permitir el intercambio de información sobre la denuncia mediante la retroalimentación de esta por parte de otros denunciantes, como la inclusión de nuevas evidencias, la adicción de hilos de comentarios, entre otros con el objetivo de ir soportando aún más la denuncia realizada. Además, se debe contemplar el uso de la tecnología blockchain para el almacenamiento de las denuncias, garantizando así, gracias a las ventajas de tecnología (almacenamiento centrado en bloques enlazados criptográficamente)

el aseguramiento de los datos almacenados, evitando la pérdida de cualquier información.

- **Seguimiento de la denuncia:** Cuando una denuncia es recibida, la entidad debe contar con un grupo de personas especialmente asignadas para gestionarla, con la formación necesaria y las competencias para tratar casos éticamente sensibles. Ellos son los encargados, además de agrupar denuncias relacionadas, de decidir el alcance de la denuncia, si se debe escalar o si se requieren expertos externos para la investigación.
- **Gestión de la denuncia:** Como parte del servicio de denuncias, se debe contar con una herramienta que permita un manejo eficaz y correcto de cada denuncia recibida; dejando en evidencia los análisis realizados y los resultados obtenidos de los mismos. Con el objetivo final de tomar las medidas apropiadas garantizando la seguridad de los datos y la privacidad.

Además, se plantea el uso de un sistema de semáforo que permita conocer el estado de una denuncia, informando las fechas correspondientes con esto. El denunciante, la ciudadanía y cualquier interesado, podrá conocer el estado y resultado de las denuncias realizadas sobre un proceso de contratación.

- **Generación de reporte del análisis/investigación realizada:** Una vez terminado el análisis/investigación de una denuncia recibida, los datos resultantes del mismo deben ser recopilados y almacenados en la misma aplicación, generando así un reporte con sus debidos soportes, este mismo puede o no ser reportado al denunciante, pero debe, en caso de ser necesario, ser reportado a las entidades gubernamentales correspondientes. Así mismo una estrategia de publicación sobre el resultado de la denuncia, haciendo uso de las ventajas de colaboratividad de la web 2.0, se debe contemplar en donde el mismo sea de conocimiento público, esto mediante un comunicado a la opinión pública, una lista de difusión masiva o una circular publicada en páginas gubernamentales.

Por otro lado, una estrategia de generación de información estadística sobre las denuncias debe ser implementada, en donde la opinión pública, pueda conocer, entre otras, las convocatorias, entidades o participantes con más denuncias recibidas.

### 3.4.3 Salida

El objetivo de este servicio, a diferencia de los anteriores, no es atacar un delito en específico que pueda presentarse en un proceso de contratación, sino recopilar denuncias ciudadanas sobre anomalías presentadas en el proceso, por lo cual, la salida del mismo, no será un indicador porcentual de la probabilidad de la presencia de un delito, sino más bien un reporte donde se recopilen aquellas denuncias recibidas sobre un proceso, además de la gestión realizada por la entidad sobre las mismas.

Este reporte servirá además para alertar a los funcionarios pertinentes sobre anomalías presentadas, lo cual será de gran utilidad para lograr detectar delitos que no son atacados de forma directa por un servicio especializado en este modelo.

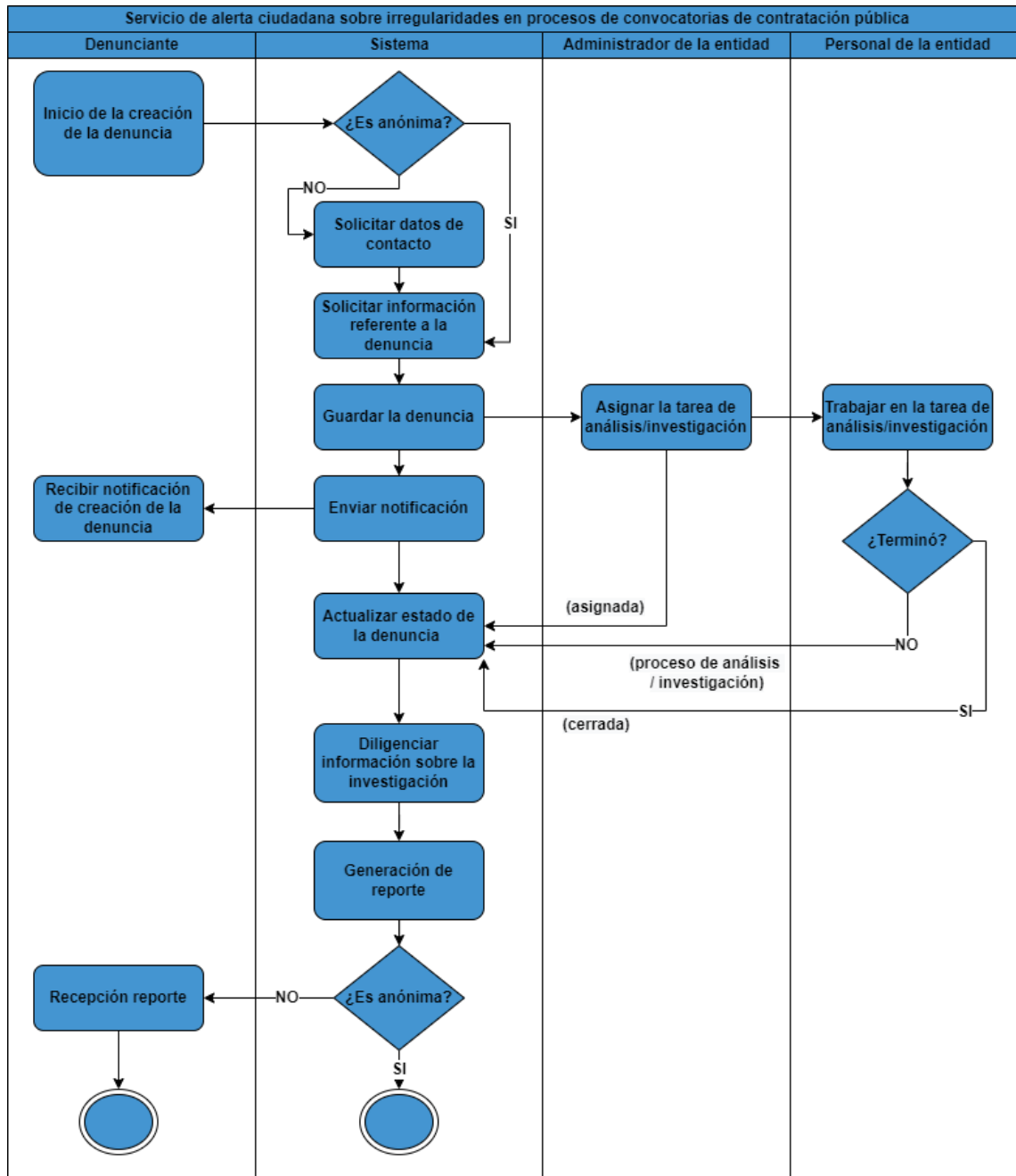


Figura 11. Diagrama de procesos de negocio servicio de denuncia ciudadana

### 3.5 Servicios transversales al modelo

El modelo planteado en esta investigación consta, como se ha explicado anteriormente, de un conjunto de servicios especializados, los cuales tienen unos objetivos específicos de funcionamiento, y que, al ser implementados, deberán poder ser accedidos y administrados por usuarios, con ello, nace la necesidad, como en cualquier aplicación software de contar primero con un servicio de administración de usuarios, el cual se encargue de gestionar el acceso y

administración de usuarios y los permisos de estos sobre las funcionalidades de la aplicación. Y segundo de un servicio administrador encargado de gestionar las variables, atributos y características administrativas de los diferentes servicios del modelo a implementar.

Estos servicios son transversales al modelo, debido a que la lógica implementada en ellos abarca a todos los servicios especializados existentes y que se deseen incluir posteriormente. A continuación, se presenta en detalle los dos servicios transversales incluidos en el modelo propuesto.

### **3.5.1 Administración, autenticación y autorización de usuarios**

Como cualquier servicio que se presenta como una aplicación web, se hace necesario implementar un módulo de administración, autenticación y autorización de usuarios, que cuente con las características típicas de este tipo de módulos. Es preciso mencionar que este módulo es independiente y transversal a todos los servicios que se deseen implementar en el modelo macro de definición de presuntas situaciones de corrupción, siendo así, este módulo puede y fue visto como un servicio complementario, que, si bien su propósito no es el de aportar directamente en el objetivo principal del modelo, si es una herramienta fundamental para este, razón por la cual no se desglosa a tan bajo nivel como los demás servicios.

Siguiendo los lineamientos encontrados en el estado del arte con respecto a módulos de control, administración y gestión de usuario [76], se logró obtener una lista de requerimientos funcionales los cuales abarcan desde un servicio de registro y acceso a la aplicación, hasta un control de permisos sobre las funcionalidades de la aplicación, esto con el objetivo de garantizar un acceso controlado a esta. Los requerimientos funcionales pertenecientes a este componente (y en un sentido más amplio, el servicio de administración, autenticación y autorización de usuarios) son:

1. Funcionalidad de control de acceso (verificación de credenciales e ingreso al servicio).
2. Funcionalidad de recuperación de acceso a la aplicación.
3. Administración de usuarios, es decir, creación (create/insert), recuperación (recovery/select), actualización (update) y borrado (delete) (CRUD) de usuarios.
4. Administración de roles y permisos de usuarios. Los permisos se asignan a los roles y estos a su vez a los usuarios de acuerdo a las funcionalidades disponibles en la aplicación.

También se contemplan una serie de requerimientos no funcionales, los cuales permiten juzgar la operatividad de la aplicación y que, de alguna manera, muchos de ellos, deben ser transversales en todos los módulos, a saber:

1. El servicio debe estar disponible en todo momento.
2. El servicio debe ser eficiente y capaz de procesar varias transacciones al tiempo, permitiendo la conexión de varios usuarios a la vez.
3. El servicio debe garantizar que el ajuste de los permisos de acceso solo puede ser realizado por parte de los administradores.

4. Debe contar con una base de datos que respalde la información de producción.
5. Las comunicaciones con servicios externos deben estar encriptadas (tanto en el canal de comunicación como los datos que se transportan).
6. Se debe garantizar un tiempo leve de aprendizaje de la aplicación por parte del usuario.
7. Se debe otorgar manuales de usuario para el fácil entendimiento de las funcionalidades del servicio.
8. Se debe proveer al usuario final, mensajes de los errores que sean informativos, claros, descriptivos y orientados a la solución de los errores, por ejemplo, en la entrada de datos.
9. La aplicación debe ser responsive, con el fin de poder garantizar la libre visualización de esta en diferentes tipos de pantallas.

De manera general, todo sistema de información, y más aún desplegado en una aplicación de tipo web, debe contar con un servicio de administración, autenticación y autorización de usuarios, que permita limitar y controlar las tareas que un usuario pueda llevar a cabo sobre el programa. En la actualidad, y gracias a los framework o marcos de trabajo, son muchas las estrategias implementadas para llevar a cabo esta tarea, facilitando así el desarrollo de la misma, puesto que son funcionalidades ya implementadas y de libre uso. La tarea entonces consiste en analizar y seleccionar cuál de los frameworks existentes es el más apropiado a la aplicación que se desea implementar, para ello, es de gran ayuda la lista de requerimientos realizada previamente, dado que el marco de trabajo escogido debe garantizar el cumplimiento a cabalidad de cada uno de los ítems listados. Dado que este módulo, es en realidad un servicio más del modelo macro, se recomienda, más no se limita, que el marco de trabajo escogido esté implementado en el mismo lenguaje de programación (o el más usado) de la aplicación y los microservicios que se van a desarrollar, con el objetivo de facilitar la comunicación entre los microservicios.

Mediante el análisis y diseño se transforman los requerimientos identificados previamente en soluciones prácticas. Para ello se debe seguir un plan que debe tener en cuenta las especificaciones y el aprovechamiento de la documentación previamente recolectada. En este proceso, se desglosó, mediante diferentes diagramas el comportamiento de cada componente previamente identificado, para este caso, el servicio de administración, autenticación y autorización de usuarios.

Inicialmente, para entender el funcionamiento que se logra con el control de acceso de usuarios, el cual se basa en roles y permisos (Role-Based Access Control, RBAC [77]) se presenta la **Figura 12**. Esta figura presenta el funcionamiento del servicio, en este se logra apreciar cómo se parte de un usuario administrador quien es el encargado de la creación inicial de los demás usuarios y segundo de la asignación de los roles que tendrán cada uno de ellos, partiendo de la idea de que cada usuario puede tener uno o más roles en la organización, y a su vez a cada rol se le asigna una serie de permisos, con los cuales se controlan el acceso de las diferentes funcionalidades disponibles. Con este diagrama, se puede entender el comportamiento que debe tener los requerimientos funcionales de administración de usuarios y administración de roles y permisos.



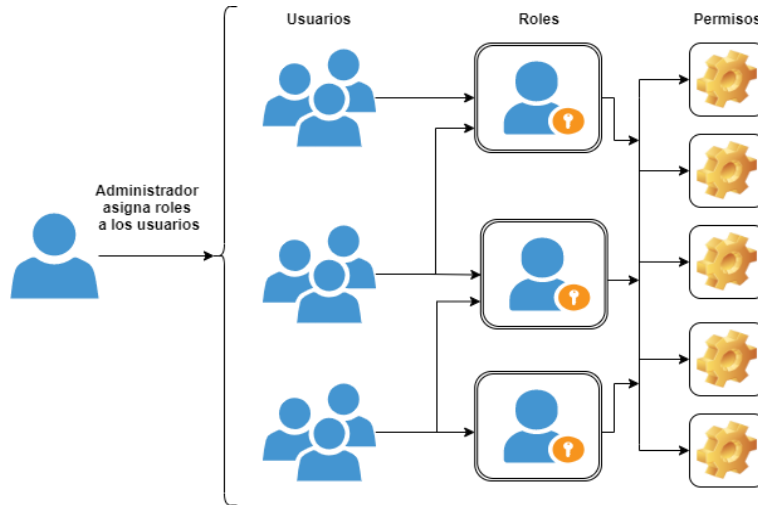


Figura 12. Arquitectura control de acceso basado en roles

Por otro lado, la **Figura 13** muestra el flujo de operaciones e interacciones de las actividades en las que se basa el servicio de administración, autenticación y autorización de usuarios, con todos los requerimientos previamente listados. En este apartado no se incluyen otros diagramas que expliquen más en detalle el servicio de administración, autenticación y autorización de usuarios dado que, como se mencionó anteriormente, su implementación se hace mediante el uso de alguna librería o framework con el objetivo de evitar el desarrollo de este.

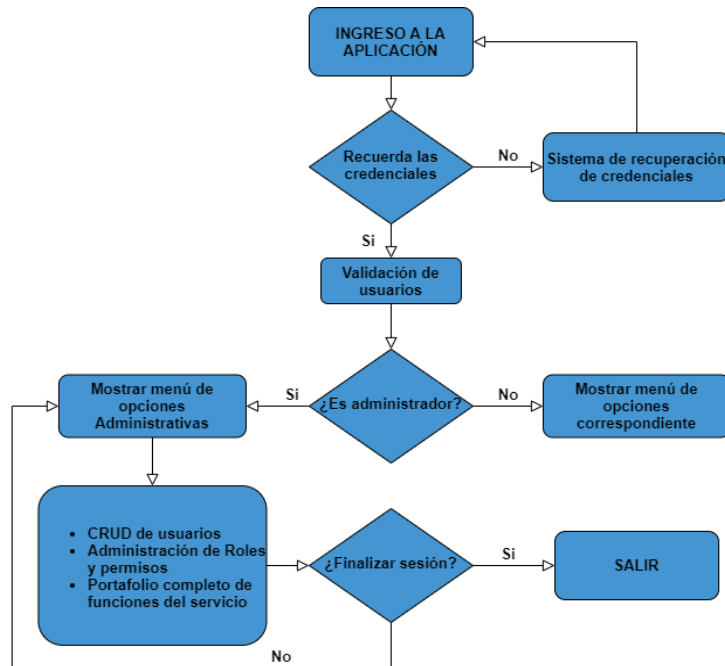


Figura 13. Flujo de administración, autenticación y autorización de usuarios

### 3.5.2 Administración de los servicios

Similar al anterior servicio, que es transversal al modelo, es necesario, como en cualquier aplicación, poder ofrecer un portafolio de configuraciones sobre la aplicación y los microservicios que usa, con el objetivo de poder tener una correcta administración del software. Este listado de operaciones, al ser de tipo administrativas, solo están dirigidas a los roles que cuenten con los permisos correspondientes. Con lo anterior, y similarmente al servicio de administración, autenticación y autorización de usuarios, este módulo se puede ver como un servicio independiente y transversal dado que también debe estar disponible para todos los servicios implementados en el modelo macro, con la salvedad que, a medida que se vayan adicionando servicios este componente irá aumentando en sus funcionalidades, dado que habrá nuevos ítems que se deben administrar. Sin embargo, gracias al planteamiento de este servicio transversal en un futuro, cuando se desee adicionar un servicio que no se haya incluido en esta tesis, se contará con una guía de cómo incluir nuevas funcionalidades administrativas que controlen el nuevo servicio. Siguiendo los mismos lineamientos mencionados en el servicio de administración, autenticación y autorización de usuarios [76], se cuenta con un portafolio de operaciones administrativas (excluyendo las de usuario) que deben ser implementadas, las cuales son:

1. Administración de la entidad (Editar la información registrada en la entidad).
2. Administración de los proponentes (CRUD de proponentes).
3. Administración de los servicios (CRUD de convocatorias).
4. Administrador de roles y usuarios propios de cada entidad (CRUD de usuarios y roles de entidad).

Por parte de los requerimientos no funcionales del componente, se pueden adaptar los listados en el anterior servicio transversal. De manera similar al anterior servicio, una vez registrada una entidad en la aplicación, la información inicial es almacenada, dicha información corresponde a la otorgada por la entidad en el momento del registro a la aplicación, sin embargo la misma puede ser modificada en cualquier momento (con excepción de algunos datos), en ese mismo sentido, y como se ha explicado anteriormente, los servicios especializados están enfocados a convocatorias de contratación estatal en donde participan proponentes, por lo cual es necesario contar con un servicio administrador tanto de convocatorias como de proponentes, que permita realizar las operaciones básicas de software (CRUD). Para ello, vuelve a ser de gran utilidad la implementación de frameworks de desarrollo, los cuales facilitan, y en muchos casos brindan el desarrollo de dichas operaciones básicas, manejando una arquitectura basada en el patrón denominado modelo-vista-controlador (Model-View-Controller, MVC) [78] con el objetivo de independizar las diferentes capas de desarrollo.

La implementación de este componente, usando frameworks de desarrollo, no conlleva mayor tiempo, debido a que estos últimos cuentan con implementaciones que facilitan el desarrollo. La **Figura 14** muestra un diagrama de secuencia que permite entender en un alto nivel la comunicación entre las diferentes capas del

desarrollo al momento de recibir una petición, en este caso, de tipo CRUD. La figura permite apreciar cómo, la petición parte del usuario desde la vista, en donde se realizan validaciones iniciales como por ejemplo tipo de datos, llenado de campos obligatorios, entre otros, luego esa petición se envía a la capa del controlador, en donde, se realizan operaciones de la lógica del negocio, para con ello continuar el viaje de la petición a la última capa, la del modelo, donde la petición, es adaptada al modelo de base de datos, para su posterior envío al gestor de base de datos y retornar la respuesta del mismo hasta hacerla llegar al usuario final. Los desarrollos de las funcionalidades de este componente siguen la estructura de este diagrama.

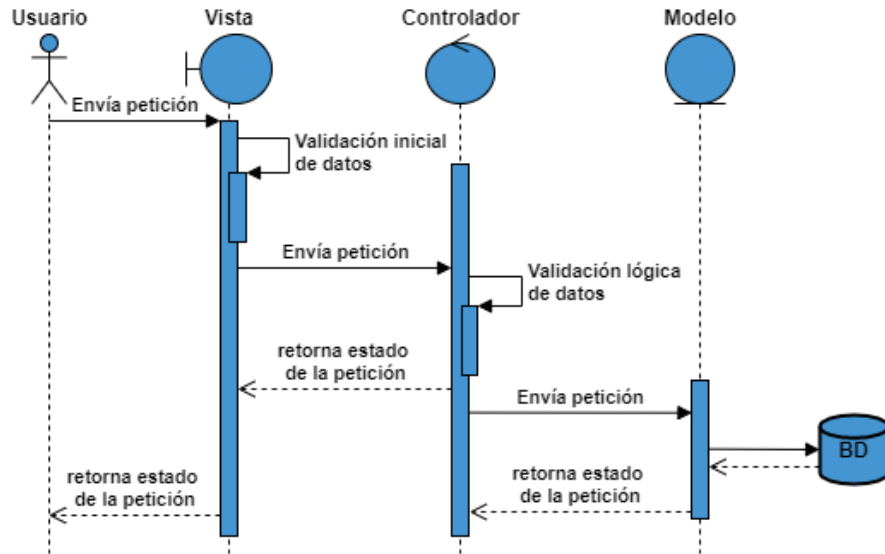


Figura 14. Diagrama de secuencia que ejemplifica una petición en MVC

Esta página ha sido intencionalmente dejada en blanco.

## CAPÍTULO 4

---

### 4. Servicios desarrollados

Al desarrollar un servicio software una interrogante inicial es la arquitectura que soportará el servicio, para este caso, se optó por hacer uso de una arquitectura de nube, con el despliegue mediante el uso de aplicaciones web (basado en la tendencia que se puede apreciar en la literatura experta a la hora de implementar sistemas de información), dadas las ventajas que esta trae: fácil despliegue en diferentes tipos de sistemas operativos, interfaz amigable y fácil de usar, más fácil de diseñar en forma escalable, de fácil actualización y centralización de la seguridad [79]. Y como característica principal, la opción de escalabilidad de los recursos a utilizar, dado que inicialmente el servicio (y la aplicación en general, que acogerá el modelo planteado) no necesitará una gama amplia de recursos computacionales por la poca demanda inicial, pero luego de un tiempo, con el aumento del uso del mismo por diferentes entidades, se estima que será necesario ampliar los recursos, característica que un servicio de **cloud computing** permite de manera fácil sin afectar el rendimiento de la aplicación.

Los servicios planteados en el presente modelo, tal y como se ha mencionado anteriormente, están dirigidos a entidades gubernamentales, que ofrezcan públicamente procesos de licitación. A su vez, se tendrá acceso a dichos servicios y la información de los mismos mediante usuarios previamente registrados en la plataforma. Es por esto por lo que, la arquitectura debe permitir la conexión de múltiples usuarios de diferentes entidades a una única instancia del servicio que maneja la información de todas las entidades con la debida separación de datos y seguridad. Esta característica se conoce como arquitectura **multi-tenant (multi-propietario)** y se presentan como una de las más utilizadas para este propósito [80]. Esta arquitectura hace referencia a uno de los principios de la arquitectura de software donde solo se ejecuta una única instancia en un servidor y múltiples clientes acceden a ella, así un solo desarrollo software satisface las necesidades de múltiples clientes, garantizando la separación de la información sensible de cada uno, así mismo esta arquitectura permite la personalización de elementos para cada cliente final.

Para el almacenamiento de los datos en esta arquitectura multi-propietario se usa una única base de datos con esquemas compartidos [80], donde cada tabla del modelo relacional maneja la identificación del cliente que es dueño de cada registro o fila de datos almacenado. Algunas de las ventajas con las que cuenta este esquema son los bajos costos de los recursos computacionales y la masificación del número de clientes permitidos en comparación con otros enfoques **multi-tenant**.

Teniendo en cuenta que la aplicación desarrollada para soportar los servicios especializados planteados en el modelo debe permitir la escalabilidad en su desarrollo debido a la futura incrustación de nuevos servicios, es necesario plantear una arquitectura que permita la creación y comunicación entre los servicios garantizando la separación de sus funcionalidades y evitando la repetición de funciones y/o procedimientos (código).

Planteado la anterior, la arquitectura escogida para el desarrollo del software, es una arquitectura basada en microservicios y microfrontends, arquitectura que es diametralmente opuesta a la ya muy usada arquitectura monolítica (una arquitectura donde el software se implementa en un solo desarrollo de manera que todos los componentes y módulos de la aplicación quedan acoplados entre sí, lo que genera un problema de ausencia de escalabilidad de la aplicación). Con los microservicios [81], se garantiza una fácil escalabilidad de la aplicación global, permitiendo además la realización de ajustes de manera independiente sin afectar mayormente el todo. Así mientras que en la arquitectura monolítica se cuenta con una única unidad, con los microservicios se tiene un conjunto de pequeños servicios, ejecutados de forma independiente y autónoma, tema que se ajusta al modelo planteado en esta tesis.

Los microservicios, nacen con el propósito inicial de suplir las necesidades y fallas del monolito del lado del servidor, o comúnmente conocido en el mundo del desarrollo de software como **backend**. Con el objetivo de suplir esto mismo del lado del cliente o **frontend** y extendido de la idea de los microservicios, se implementa una arquitectura de microfrontends [82] aplicando la tendencia actual del desarrollo de software de las aplicaciones **single page app**. El objetivo principal de los microfrontends es tratar una aplicación web como un conjunto de características independientes cada una con su propia lógica de negocio que es justamente lo ofrecido, del lado del servidor, por los microservicios. Con esto, el modelo implementado hace uso de estas dos arquitecturas específicas permitiendo la comunicación cliente-servidor mediante el ofrecimiento de APIs [81] desde el lado del servidor.

El desarrollo del software que soporta tanto este servicio como el modelo como tal se realizó siguiendo una metodología de desarrollo de software, y usando patrones de diseño recomendados en la literatura [83], haciendo uso de lenguajes de programación idóneos tanto para el backend como el frontend. Los lenguajes escogidos para cada capa del servicio cumplen con las siguientes características mínimas:

- Del lado del backend se cuenta con un framework que facilitó el desarrollo del control y administración de usuarios (buscado evitar el desarrollo de todo o parte de este módulo). Así mismo cuenta con frameworks que facilitaron el desarrollo, despliegue y publicación de APIs para su posterior uso. Finalmente, cuenta con una amplia gama de librerías que ahorran el desarrollo de funcionalidades como lectura de archivos, conversión de texto, entre otras. El lenguaje escogido para esta capa fue Python en su versión 3.9.0.

- Del lado del frontend cuenta con plantillas html que permitan su fácil uso para el despliegue del servicio de forma responsiva (también conocida como adaptativa, que hace referencia a la capacidad de que una misma página html se vea correctamente en distintos dispositivos), así como la opción de implementar una arquitectura de módulos para la independencia de los servicios. El lenguaje escogido para esta capa fue TypeScript el cual es un superconjunto de JavaScript, y específicamente se usó el framework Angular en su versión 11.

A continuación, se desglosa el desarrollo realizado para implementar dos de los servicios especializados propuestos en esta investigación.

#### 4.1 Desarrollo del servicio de detección de similitud en propuestas técnicas

El servicio es distribuido y escalable, y da acceso a diferentes entidades estatales que ofertan convocatorias de contratación. Al software desarrollado para soportar el modelo y, en este caso, el servicio de detección de similitudes, acceden los usuarios de cada entidad y pueden consultar/operar información referente a su respectiva entidad, como las convocatorias realizadas, los ofertantes, entre otros. En específico, para este servicio, al momento de realizar un análisis de similitud, la aplicación permite tres opciones principales:

1. Un análisis general sobre toda la base de datos, es decir, todas las propuestas técnicas registrados sin importar la entidad de origen desde que estos sean de acceso público,
2. Un análisis de similitud filtrado sobre las propuestas técnicas registradas por la misma entidad y,
3. Un análisis de similitud filtrado sobre las propuestas de una convocatoria específica a la que pertenece la propuesta técnica que se va a analizar.

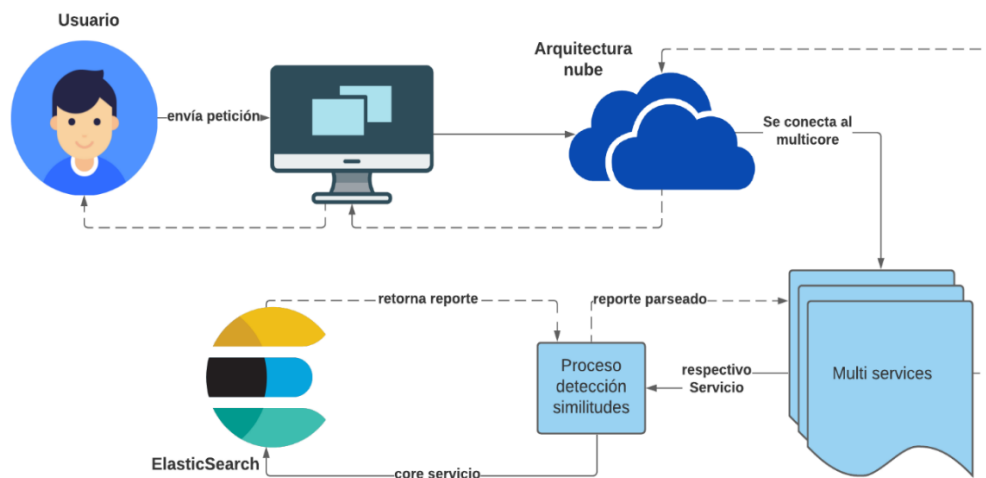


Figura 15. Diagrama general del servicio de similitud en propuestas técnicas

La **Figura 15** presenta los componentes generales del servicio de detección de similitudes, en donde se identifica que el proceso parte del envío de una petición al servicio por parte del usuario, para los cuales se contempla los siguientes roles:

1. **Administrador:** Es el rol encargado de la administración y mantenimiento general del servicio.
2. **Súper usuario (root):** Es un rol asociado a cada una de las entidades registradas en el servicio y se encarga de asignar los demás usuarios de su propia entidad de acuerdo con sus roles (root, archivador, analista), entre otras cosas.
3. **Archivador:** Es un rol asociado a cada una de las entidades registradas en el servicio y se encarga del almacenamiento de las propuestas técnicas asociadas a cada una de las convocatorias de contratación, por ejemplo: términos de referencia, decretos, reglamentos, directrices, anexos, propuestas, evaluaciones, actas y demás. Al registrar las propuestas técnicas principales de cada convocatoria define que partes del documento son repetibles (por ejemplo, cuando se usan pliegos únicos o plantillas de presentación de las propuestas) en las diferentes propuestas presentadas por los proveedores (o contratistas), es decir, no se deben tener en cuenta como similares en el proceso de evaluación de similitud. Este rol además es el encargado de definir y modificar el carácter público o privado de las propuestas que se suben al servicio.
4. **Analista:** Es un rol asociado a cada una de las entidades registradas en el servicio y se encarga de evaluar cada propuesta presentada en la convocatoria frente a las propuestas técnicas disponibles en todo el servicio (análisis global) o frente a las propuestas técnicas de la misma entidad (análisis local) o frente a las propuestas técnicas de la misma convocatoria (análisis por convocatoria). Con el reporte generado por el servicio toma decisiones sobre cada propuesta. Este usuario tiene la opción de conocer estadísticas y reportes de tendencias de los documentos de su propia entidad. Los reportes de análisis destacan uno a uno los textos (párrafos o frases) que son iguales o parcialmente similares con otras propuestas técnicas, definiendo cuáles son esas propuestas técnicas y reportando un porcentaje de similitud contra cada propuesta técnica comparada y el valor de similitud obtenido por la propuesta técnica que se analiza. El proceso obvia los textos que se han marcado como repetibles por parte del analista.
5. **Buscador:** Es el rol que permite a usuarios anónimos consultar los documentos públicos (términos de referencia, propuestas, contratos, entre otros) de las convocatorias disponibles en el servicio. Estos documentos pueden pertenecer a cualquier entidad registrada y a cualquier convocatoria realizada por dichas entidades. Este usuario también tiene la opción de conocer estadísticas generales y reportes de tendencias de los documentos públicos. La funcionalidad disponible para este rol de usuario, por defecto está integrada en las funcionalidades de los demás roles.

Como característica adicional, el servicio permite la generación de información estadística sobre la herramienta en sí, como por ejemplo datos generales sobre los análisis de similitudes realizados, cantidad de estos, entre otros, estos análisis



pueden ser usados posteriormente por usuarios externos para un análisis profundo de la situación de convocatorias estatales, y plantear algunas estrategias. Es válido recalcar que esta característica adicional solo se fundamenta en aquella información que es de carácter pública y en ningún momento se pondrá en evidencia información sensible o específica de una entidad o sus documentos.

El objetivo principal de este servicio es detectar similitud entre una propuesta técnica nueva y una fuente de propuestas almacenadas (global, por entidad o por convocatoria), con esto se hace necesario contar, primero, con una estrategia de almacenamiento de información textual. En la literatura, una estrategia recomendada, es hacer uso de técnicas de recuperación de la información como la generación de índices invertidos [47], en donde el documento no es almacenado de manera directa enlazándolo con un identificador, sino que el mismo es desglosado por palabras (tokens que pueden ser no solo palabras, sino n-gramas o frases) y en cada una de ellas se apunta la referencia de los documentos de la fuente de información en donde la palabra aparece. Esta técnica es altamente recomendada y utilizada en la literatura para este tipo de problemas dado que permite la realización de búsquedas complejas, en tiempo relativamente bajo. Ahora bien, existen algunas librerías o motores de búsqueda de libre uso que cuentan con un amplio pool de funciones para soportar esta tarea. Para el desarrollo del core del almacenamiento y búsqueda de información en esta investigación, se hace uso de este tipo de estrategias, dado que simplifican el desarrollo software y ofrecen una variedad de funcionalidades que son de gran utilidad para el servicio de detección de similitudes, específicamente, en este desarrollo se usó la librería Natural Language Toolkit NLTK (<https://www.nltk.org/>).

Finalmente, a continuación, se presentan una serie de pantallas correspondientes a algunas de las funciones principales del desarrollo del servicio de detección de similitudes entre propuestas técnicas.

La **Figura 16** muestra una pantalla inicial donde se logra apreciar, en primer lugar, la estructura general del aplicativo, el cual cuenta, y siguiendo con los lineamientos de plantillas usadas en aplicaciones web de sistemas de información, con un menú de opciones en la barra lateral izquierda, y un cuerpo en donde se desglosan las opciones a realizar, en específico la pantalla muestra el listado de convocatorias publicadas por la entidad, convocatorias a las cuales en un posterior paso se subirán los documentos de propuestas técnicas y económicas las cuales son la fuente de información para los servicios de detección de similitudes y detección de manipulación de ofertas respectivamente.

En la pantalla mostrada por la **Figura 17** se listan las propuestas técnicas subidas a la plataforma, junto con las acciones que se pueden realizar sobre ellas, el primer bloque de acciones corresponde a las opciones de visualización del documento, edición y borrado del mismo, finalmente las acciones correspondientes a análisis permiten, la ejecución del análisis de similitud, la indexación en el motor **ElasticSearch** y por último la visualización de los análisis de similitud realizados a las propuestas técnicas.

Id	Nombre	Descripción	Fecha inicio	Fecha fin	Opciones
6	IDECUT IPMC 005-2022	PRESTACIÓN DE SERVICIOS LOGÍSTICOS PARA GARANTIZAR LA PARTICIPACIÓN DE REPRESENTANTES DEL SECTOR TURISMO EN EL XI ENCUENTRO NACIONAL DE TURISMO COMUNITARIO	04/24/2022, 17:34:00	05/31/2022, 17:34:00	[Edit] [Delete] [Refresh]
5	CMA-515-2022 (Presentación de oferta)	REALIZAR LA FORMULACIÓN DEL PLAN DE ORDENAMIENTO DEL RECURSO HÍDRICO DE CAÑOS HUMACHICA Y GRANDE EN LOS MUNICIPIOS DE GUAMAL Y CASTILLA LA NUEVA EN EL DEPARTAMENTO DEL META JURISDICCIÓN DE CORMACARENA	04/29/2022, 17:33:00	05/30/2022, 17:33:00	[Edit] [Delete] [Refresh]
4	CD-UM-001-2022	OBRAS DE MITIGACIÓN PARA LA ADECUACION DE LA RED DE CONDUCCION PARA EL SERVICIO DE AGUA POTABLE EN EL MUNICIPIO DE LA PINTADA	05/01/2022, 00:00:00	05/31/2022, 23:59:00	[Edit] [Delete] [Refresh]
3	C1 - SERVICIO APOYO MANTENIMIENTO RUTINARIO	SERVICIO APOYO MANTENIMIENTO RUTINARIO EMPRENDEDORES RURALES EN EL CORREDOR FÉRREO ZIPACÓN-CACHIPAY PK 48+500 AL PK 57-000	05/06/2022, 14:18:00	05/20/2022, 14:18:00	[Edit] [Delete] [Refresh]
2	IDECUT IPMC 005-2022	PRESTACIÓN DE SERVICIOS LOGÍSTICOS PARA GARANTIZAR LA PARTICIPACIÓN DE REPRESENTANTES DEL SECTOR TURISMO EN EL XI ENCUENTRO NACIONAL DE TURISMO COMUNITARIO	2022-04-01T10:29	2022-04-30T10:29	[Edit] [Delete] [Refresh]
1	ENTREGA DEL COMPLEMENTO DEL PROGRAMA DE ALIMENTACIÓN ESCOLAR (PAE).	ENTREGA DEL COMPLEMENTO DEL PROGRAMA DE ALIMENTACIÓN ESCOLAR (PAE), A TRAVÉS DEL CUAL SE BRINDA RACIONES ALIMENTARIAS INDUSTRIALIZADA Y PREPARADA EN SITIO A LOS NIÑOS, NIÑAS Y ADOLESCENTES DE LA MATRICULA OFICIAL DEL MUNICIPIO DE PUERTO TRIUNFO-ANTIOQUIA, DE CONFORMIDAD CON LOS LINEAMIENTOS TÉCNICO ADMINISTRATIVOS DEL PROGRAMA DE ALIMENTACIÓN ESCOLAR DEL MINISTERIO DE EDUCACIÓN NACIONAL	09/01/2021, 12:23:00	09/30/2021, 12:23:00	[Edit] [Delete] [Refresh]

Figura 16. Pantalla: Listado de convocatorias

Title	Descripción	Convocatoria	Opciones	Análisis
Propuesta técnica ÉBANO DC S.A.S 5922521f-02a5-4f09-8628-eda6e7dca73	Propuesta técnica ÉBANO DC S.A.S	IDECUT IPMC 005-2022	[Edit] [Delete] [Refresh]	[Análisis]
Propuesta técnica Fungestores f9f0d2a4-0535-4f7f-883b-aaba6e5d9a40	Propuesta técnica Fungestores	IDECUT IPMC 005-2022	[Edit] [Delete] [Refresh]	[Análisis]
Propuesta técnica GLC INVERSIONES SAS 25fe30e5-f712-46d2-8286-70ee250a263b	Propuesta técnica GLC INVERSIONES SAS	IDECUT IPMC 005-2022	[Edit] [Delete] [Refresh]	[Análisis]
Propuesta técnica HUMACHICA - GRANDE 0c72b364-b398-44a8-99ef-7aeb58771691	Propuesta técnica HUMACHICA - GRANDE	IDECUT IPMC 005-2022	[Edit] [Delete] [Refresh]	[Análisis]

Elementos por página 5 - 1 - 4 of 4

Figura 17. Pantalla: Listado de propuestas técnicas

Finalmente, la pantalla ilustrada en la **Figura 18** se muestra el análisis de similitud generado para un documento en específico. El análisis consta de 3 partes, en la primera se muestran estadísticas generales relacionadas al documento y al análisis, como los son, la cantidad de frases que tiene el documento, el número de frases en donde el porcentaje de similitud supera un umbral fijado previamente y el porcentaje de similitud del documento con respecto a la fuente de información. En la segunda parte del análisis se listan las frases que superaron el umbral fijado previamente, así como la frase del documento con el que más se presentó similitud, indicando, con una paleta de colores las palabras o tokens que son similares (Rojo: copiado textualmente, Naranja: parafraseo de la palabra y Amarillo: palabras ligeramente relacionadas), el reporte permite, de ser necesario, marcar una frase como “frase común” con el objetivo de excluirla del presente y de futuros análisis realizados. Por último, el reporte presenta un listado de documentos que presentaron similitud con el analizado, con el propósito de realizar un análisis uno a uno.

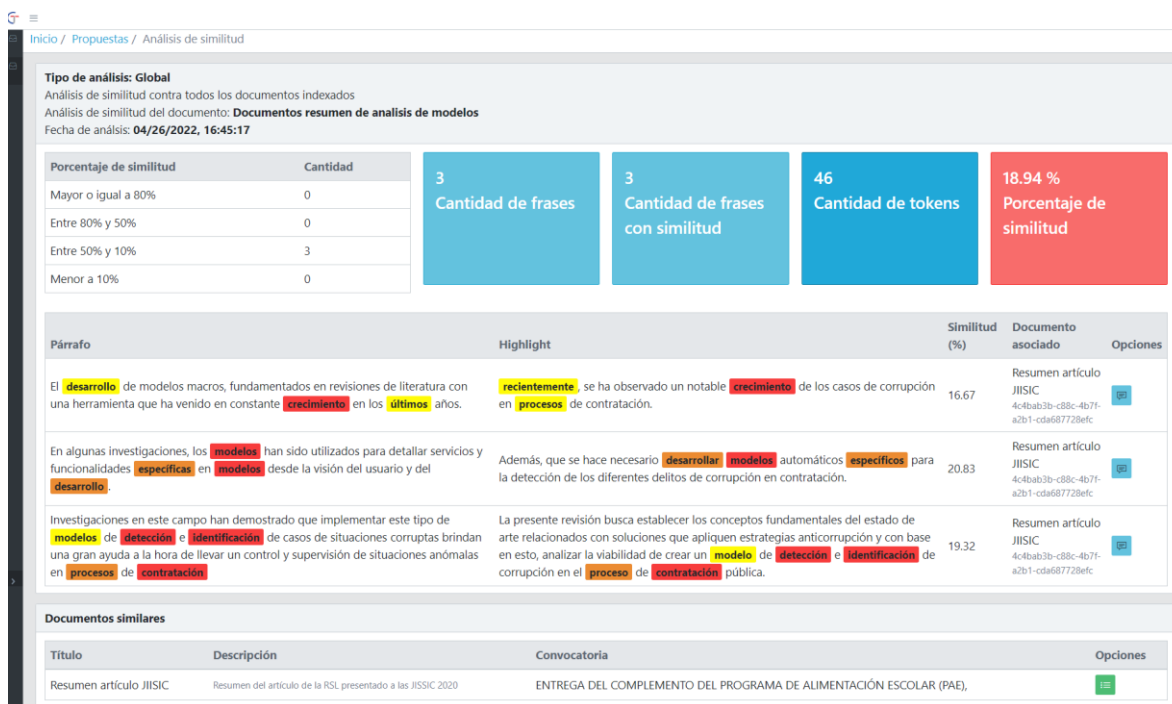


Figura 18. Pantalla: Resultado del análisis de similitud

## 4.2 Desarrollo del servicio de detección de probable manipulación de ofertas

El servicio cuenta con un sistema de acceso a usuarios pertenecientes a las entidades que accedan a la solución como un todo (servicio transversal del modelo propuesto). Al momento de realizar un análisis con este servicio, este se realiza sobre todas las propuestas económicas relacionadas con una misma convocatoria.

En relación con los roles y permisos, este servicio cuenta con los mismos mencionados en el servicio de detección de similitudes con la salvedad que el rol de **Archivador** está enfocado a estandarizar el contenido de las propuestas técnicas de manera que el servicio pueda operar de manera correcta, así mismo el rol de **Analista** se encarga de inspeccionar y analizar el reporte generado por el servicio sobre la convocatoria evaluada, tomando las decisiones pertinentes sobre la misma, además cuenta con la opción de conocer estadísticas y reportes relacionadas con las propuestas técnicas y convocatorias evaluadas.

La principal funcionalidad de este servicio es detectar la probabilidad de la presencia del delito de manipulación de ofertas entre las propuestas económicas relacionadas con una convocatoria, por esto, similar al anterior servicio se debe contar con una funcionalidad de extracción y almacenamiento de los datos que se encuentran registrados en el documento, para ello se siguió una estrategia basada en técnicas de aprendizaje automático supervisado [53], el resultado de esta extracción se almacena en una base de datos NoSQL aprovechando las ventajas de su bajo costo y celeridad de almacenamiento y estructura.

Seguidamente, se inicia la detección de probabilidad de manipulación de ofertas, para esto, se siguió una estrategia de combinación de uso de pantallas estadísticas propuestas por los autores en [84] (coeficiente de variación, curtosis, diferencias porcentuales y relativas normalizadas) con las cuales se logra obtener un índice de probabilidad de presencia de manipulación de ofertas sobre la convocatoria a evaluar. Además del algoritmo de LexRank planteado en la **Figura 6** que busca determinar un valor de probabilidad de pertenencia al delito de manipulación de ofertas a cada propuesta económica presentada en la convocatoria que se está analizando.

A continuación, se presentan una serie de pantallas correspondientes a las funciones principales del desarrollo del servicio de detección de manipulación de ofertas entre propuestas económicas.

En la **Figura 19** se listan las propuestas económicas subidas a la plataforma, junto con las acciones que se pueden realizar sobre ellos, correspondiente a las opciones de visualización del documento, edición y borrado del mismo. Por otro lado, las acciones de análisis se muestran al inicio de la grilla, donde se permite el cargue de nuevas propuestas económicas (Subir Propuestas), el análisis de la manipulación de ofertas sobre la convocatoria (Ejecutar análisis), y por último la visualización de los análisis realizados sobre la convocatoria (Ver análisis).



**Figura 19.** Listado de propuestas económicas de una convocatoria

En la **Figura 20** se muestra el análisis de presencia del delito de manipulación de ofertas generado para una convocatoria. Similar al servicio anterior, el análisis consta de tres partes, en la primera se muestran el cálculo de las pantallas estadísticas hechas sobre la convocatoria (coeficiente de variación, curtosis, diferencias porcentuales y relativas normalizadas). La agrupación de estas pantallas estadísticas permite la generación de un índice de probabilidad de presencia del delito dentro de la convocatoria, el cual se muestra en la figura, al finalizar la primera parte del análisis. En la segunda parte del análisis se listan los productos y/o

servicios ofertados en la convocatoria, junto con el calor de las pantallas estadísticas mencionadas anteriormente, esto con el objetivo de presentarle al usuario final los cálculos hechos para la probabilidad de presencia de delito, esta vez por producto y/o servicio. Por último, el reporte muestra el listado de ofertas económicas presentadas en la convocatoria, junto con su probabilidad de participación en el delito de manipulación de ofertas, esto se logra con el uso del algoritmo **LexRank** implementado y adaptado en esta investigación (**Figura 6**).



**Figura 20.** Resultado del análisis de manipulación de ofertas (3 propuestas partícipes del delito)

Esta página ha sido intencionalmente dejada en blanco.

## CAPÍTULO 5

---

### 5. Evaluación del modelo y los servicios desarrollados

#### 5.1 Evaluación conceptual

En este apartado del documento se presenta el proceso de evaluación realizado al modelo propuesto en esta investigación, mediante la implementación de la técnica de investigación cualitativa de Grupo Focal con el objetivo de tener una retroalimentación sobre la primera versión propuesta del modelo con un grupo de tres expertos y luego se realizó una evaluación formal al modelo y los servicios desarrollados por parte de otros tres expertos en el tema.

Como se comentó, el proceso de evaluación se dividió en dos secciones. La primera tuvo como objetivo conocer la opinión y recibir retroalimentación por parte de tres expertos (dos en el área de anticorrupción y uno desde la perspectiva de la Ingeniería de sistemas o informática) en relación con el modelo propuesto, buscando con esto obtener información relevante para mejorar el modelo propuesto, así como los servicios que lo sustentan. Para la recopilación de la información obtenida en el grupo focal se empleó los lineamientos definidos en [85], cuya estructura se compone de las siguientes fases:

1. **Planeación de la investigación:** donde se establece el contenido y el procedimiento que se aplicarán en la sesión.
2. **Definición de grupos de discusión:** donde se caracterizan e identifican los participantes y se constituyen los grupos de debate.
3. **Conducción de la sesión de debate:** donde se ejecutan los procedimientos establecidos para dar cumplimiento al debate. Esta fase es coordinada por un moderador y desarrollada por los participantes.
4. **Análisis de información y reporte de resultados:** donde se emplean métodos de análisis para la obtención de la información de valor sobre el debate.

En la realización del debate, la primera etapa se realizó una introducción y contextualización del problema a abordar en el trabajo de investigación, luego se procedió a la presentación de la solución propuesta (modelo y servicios relacionados). Esto se hizo mediante una presentación donde se mostró cada uno de los diagramas del modelo y servicios, con su respectiva explicación general y por componente. Cada vez que se iban presentando los componentes del modelo, así como los servicios, se daba un tiempo para conocer las opiniones, comentarios, observaciones y sugerencias que los participantes tuvieran. Esto se realizó, primero con una sección de participación abierta, a modo de lluvia de ideas, y segundo respondiendo una encuesta estructurada donde se planteaban unas preguntas

generales sobre lo presentado (ver **Tabla 9**). El desarrollo de la sesión de debate se definió siguiendo el protocolo descrito en la **Tabla 10**.

**Tabla 9.** Preguntas realizadas en el debate

N°	Pregunta
1	¿Quedaron claros los procesos expuestos anteriormente?
2	¿Cree usted que el servicio expuesto es claro? ¿Las conexiones son claras?
3	¿El servicio es lo suficientemente detallado?
4	¿Considera que las entradas de los servicios son las adecuadas?
5	¿Qué sugerencias aportaría para mejorar el servicio expuesto?
6	¿Es entendible la conexión entre el servicio expuesto y su relación con el modelo?

**Tabla 10.** Protocolo de la sesión del Grupo Focal

<b>Tema</b>	Modelo para definir índices de riesgo de presuntas situaciones de corrupción en convocatorias de contratación estatal en Colombia basada en Big Data y Procesamiento del Lenguaje Natural
<b>Duración</b>	2 horas
<b>Método</b>	Virtual
<b>Fecha</b>	29 de abril de 2022
<b>Hora</b>	1:30 pm
<b>Moderador</b>	Ing. Julio César Luna Ortega
<b>Supervisor</b>	PhD. Carlos Alberto Cobos Lozada
<b>Objetivo general</b>	Obtener información relevante para mejorar el modelo presentado en este proyecto de investigación.
<b>Métodos de captura de información</b>	<ul style="list-style-type: none"> <li>• Archivo de video de la sesión de debate.</li> <li>• Anotaciones de comentarios, sugerencias y observaciones.</li> <li>• Formularios con las respuestas a las preguntas realizadas.</li> </ul>

Para la definición de los participantes del grupo de discusión se buscó contar con la participación de profesionales expertos en el área de Anticorrupción y de las áreas de la Ingeniería de Software, Sistemas de información e Inteligencia Artificial. El listado de los participantes se muestra en la **Tabla 11**. Por último, Se hizo un proceso de análisis de la información recolectada, estableciendo los aportes y observaciones no repetidas y clasificándolas.

**Tabla 11.** Participantes de la sesión Grupo Focal

ID	Nombre	Experiencia profesional, académica o administrativa	Ocupación
1	Adriana María Romero	MSc. en Ciencia Política y Economía Política del London School of Economics and Political Science. Sus áreas de interés son la transparencia, análisis de corrupción, rendición de cuentas y economía política del desarrollo	Directora del Área de Estudios y Política Pública del Instituto Anticorrupción
2	Ivón Sepúlveda	Comunicadora social egresada de la Universidad del Norte con experiencia en investigación académica en ciencias sociales. Ha colaborado en la realización de proyectos de investigación sobre educación inclusiva y	Asesora de análisis de datos en el Instituto Anticorrupción



		asistencia de procesos editoriales para revistas científicas	
3	Hugo Armando Ordoñez Erazo	Ingeniero de Sistemas, Magister en Computación y Doctor en Ingeniería Telemática. Investigador Senior según Minciencias en Ingeniería de Software, Ciencia de Datos, Descubrimiento de Procesos de Negocio y Gestión de la Información.	Docente Planta de la Universidad del Cauca

### 5.1.1 Resultados del grupo focal de retroalimentación

Contando con la información recolectada en el debate, se lograron identificar los aportes y observaciones realizadas por cada uno de los participantes. La **Tabla 12** sintetiza la información recolectada.

**Tabla 12.** Aspectos por mejorar identificados en la sesión del Grupo Focal

Aspectos por mejorar y observaciones
Detallar de una mejor manera la base de datos que alimenta los servicios
Poner nombres más significativos
Usar una mejor representación cuando se indique que se ejecutan actividades en paralelo
Tener en cuenta métodos para mitigar los limitantes para acceder a la información
Tener en cuenta un mecanismo de rastreo de empresas internacionales en el servicio de detección de carteles (OpenCorporates)
Tener en cuenta todas las modalidades de contratación en el servicio de detección de manipulación de ofertas
Hacer un correcto uso de los verbos a la hora de especificar las tareas
Contemplar la estrategia para extraer contenido de imágenes, para los documentos que fueron escaneados

Con la retroalimentación recibida por parte de los participantes del grupo focal, se realizaron las mejores correspondientes a cada uno de los servicios presentados, ajustando los modelos, realizando mejoras a la descripción de las entradas y salidas, mejorando la descripción de las actividades relacionadas en los procesos y definiendo nombres más apropiados para las actividades y/o procesos.

### 5.1.2 Resultados del proceso de evaluación formal

Luego de realizar las mejoras al modelo y a los servicios se realizó una evaluación formal de las nuevas versiones, contando, en este caso, con la participación de un nuevo grupo de profesionales con amplia experticia en el área de Ingeniería de Software, Sistemas de información e Inteligencia Artificial. El listado de los participantes se muestra en la **Tabla 13**. Para esta etapa de evaluación, se envió un documento completo con la caracterización de cada uno de los modelos correspondientes a los servicios propuestos, con el fin de recibir una evaluación rigurosa en aspectos referentes a su estructura, su contenido y, sobre todo, su claridad con relación a los conceptos tratados en el presente proyecto de

investigación. Para ello se proporcionó a los evaluadores, un conjunto de directrices de calidad que se debían tener en cuenta durante la evaluación de los modelos de los servicios propuestos. Para las directrices se adaptaron los lineamientos propuestos por Montes de Oca en [86], los cuales se presentan en la **Tabla 14**.

**Tabla 13.** Evaluadores de los modelos

ID	Nombre	Experiencia profesional, académica o administrativa	Ocupación
1	Fernando Gutiérrez Paramio	Ingeniero Superior de Informática (Universidad de Extremadura, España). Gerente de Desarrollo (líder) en Transformación Digital e Innovación basada en Big Data y Machine Learning.	Sr. VP Architecture en Siigo S.A.S.
2	Edgar Fabián Ruano Daza	Ingeniero de Sistemas y Magister en computación (Universidad del Cauca). Líder Técnico en desarrollo de proyectos software de alta calidad. Experiencia en investigación relacionada con optimización y procesos de negocio.	Tech Lead en Siigo S.A.S.
3	Carlos Alberto Ardila Albarracín	Ingeniero de Sistemas (Universidad Industrial de Santander) y Magister en computación (Universidad del Cauca). Experiencia en Investigación, Desarrollo e Innovación en Calidad de Software y Mejora de procesos de desarrollo de software.	Docente Planta de la Universidad del Cauca

**Tabla 14.** Lineamientos de calidad para la evaluación de los modelos

Representación visual	
<b>Diseño:</b> Aspectos relacionados con el diseño de los modelos que permite una mejor comprensión y legibilidad de estos.	<b>Etiquetado:</b> Aspectos relacionados con el etiquetado e información agregada en los elementos que representan el modelo.
<ol style="list-style-type: none"> <li>1. Minimizar el número de cruces de flujos de secuencia.</li> <li>2. Maximizar el número de flujos de secuencia dibujados perpendicularmente.</li> <li>3. Evitar solapamientos de actividades, compuertas y eventos.</li> <li>4. Evitar solapamientos entre flujos de secuencia y otros elementos del modelo.</li> <li>5. Mantener consistente la dirección de los flujos de secuencia.</li> <li>6. Mantener uniforme la longitud de los flujos de secuencia.</li> <li>7. Usar un tamaño uniforme para sus elementos.</li> <li>8. Hacer modelos largos y delgados, en lugar de cuadrados.</li> <li>9. Colocar las actividades y eventos simétricamente.</li> <li>10. Usar un estilo uniforme para la distribución de elementos y flujos de secuencia.</li> <li>11. Mantener el diagrama tan ordenado como sea posible.</li> </ol>	<ol style="list-style-type: none"> <li>1. Usar etiquetas del tipo verbo-sustantivo para las actividades.</li> <li>2. Especificar los tipos de actividades.</li> <li>3. Usar etiquetas cortas.</li> <li>4. Etiquetar los eventos.</li> <li>5. Etiquetar las actividades.</li> <li>6. Etiquetar todos los flujos de secuencia de salida.</li> <li>7. Etiquetar los contenedores de procesos.</li> <li>8. Reservar las palabras enviar/recibir para actividades y eventos de este tipo.</li> </ol>
Complejidad general	

<p><b>Morfología:</b> Son directrices prácticas cuyo objetivo es la mejora de la corrección semántica de los modelos, es decir permite la estructuración adecuada de los modelos.</p>	<p><b>Tamaño:</b> Aquellas que sugieren el número de elementos que debe tener un modelo para mantener la compresión del diagrama y minimizar los errores.</p>
<ol style="list-style-type: none"> <li>1. Para cada nodo de decisión establecer un nodo de unión del mismo tipo.</li> <li>2. No usar más de cinco flujos de secuencia en cada proceso.</li> <li>3. Utilizar a lo sumo cuatro flujos de secuencia de salida en los eventos.</li> <li>4. Evitar los ciclos en los modelos de procesos.</li> <li>5. Evitar los ciclos no estructurados.</li> <li>6. Evitar altos niveles de paralelismo.</li> <li>7. Cada camino paralelo debe sincronizarse o alcanzar un evento de fin.</li> <li>8. Evitar anidamientos profundos de bloques estructurados.</li> <li>9. Evitar el uso de elementos y fragmentos duplicados.</li> <li>10. Evitar el uso de elementos innecesarios.</li> <li>11. Mantener el camino desde el evento de inicio hasta el evento final tan corto como sea posible.</li> <li>12. Minimizar las bifurcaciones utilizadas.</li> <li>13. Seleccionar la alternativa más sencilla al modelar.</li> </ol>	<ol style="list-style-type: none"> <li>1. Evitar modelos con más de 31 elementos.</li> <li>2. Evitar modelos con más de siete eventos.</li> <li>3. No utilizar más de dos eventos de inicio.</li> <li>4. No omitir el evento de inicio.</li> <li>5. No utilizar más de dos eventos de fin.</li> <li>6. No omitir el evento de fin.</li> <li>7. Evitar modelos con un alto número de eventos intermedios.</li> <li>8. Evitar modelos con más de 34 flujos de secuencia.</li> <li>9. Evitar modelos con más de 12 procesos complejos.</li> <li>10. Evitar modelos con más de 31 actividades.</li> </ol>

Con la información suministrada por parte de los evaluadores, a continuación, se presentan los resultados de la evaluación cualitativa, a nivel conceptual, de cada uno de los servicios propuestos con su respectivo grado de calificación de acuerdo con una escala Likert de 1 a 5. Los resultados de las evaluaciones se resumen y presentan en la **Tabla 15**, la **Tabla 16**, la **Tabla 17** y la **Tabla 18**.

**Tabla 15.** Evaluación del servicio de detección de similitud de propuestas

Participantes		P1, P2, P3 (Tabla 13)		
N°	Preguntas	Totalmente de acuerdo (5) De acuerdo (4) Indeciso (3) En desacuerdo (2) Totalmente en desacuerdo (1)		
		P1	P2	P3
1	¿Considera que la descripción del servicio mantiene una secuencia lógica y clara?	5	5	5
2	¿Considera que los diagramas que representan los procesos y subprocesos cumplen con las directrices de diseño sugeridas?	5	5	5
3	¿Considera que los diagramas que representan los procesos y subprocesos cumplen con las directrices de estilo de etiquetado sugeridas?	5	3	5
4	¿Considera que los diagramas que representan los procesos y subprocesos cumplen con las directrices de morfología sugeridas?	5	3	5
5	¿Considera que los diagramas que representan los procesos y subprocesos cumplen con las directrices de tamaño sugeridas?	5	4	5

6	¿Considera que se explica y representa de manera correcta los procesos de generación del análisis?	4	3	5
<b>Promedio individual</b>		<b>4,83</b>	<b>3,83</b>	<b>5,00</b>
<b>Promedio total</b>		<b>4,55</b>		

**Tabla 16.** Evaluación del servicio de detección de manipulación de ofertas

Participantes		P1, P2, P3 (Tabla 13)		
N°	Preguntas	Totalmente de acuerdo (5) De acuerdo (4) Indeciso (3) En desacuerdo (2) Totalmente en desacuerdo (1)		
		P1	P2	P3
1	¿Considera que la descripción del servicio mantiene una secuencia lógica y clara?	5	4	5
2	¿Considera que los diagramas que representan los procesos y subprocesos cumplen con las directrices de diseño sugeridas?	5	5	5
3	¿Considera que los diagramas que representan los procesos y subprocesos cumplen con las directrices de estilo de etiquetado sugeridas?	5	3	5
4	¿Considera que los diagramas que representan los procesos y subprocesos cumplen con las directrices de morfología sugeridas?	4	3	4
5	¿Considera que los diagramas que representan los procesos y subprocesos cumplen con las directrices de tamaño sugeridas?	5	4	5
6	¿Considera que se explica y representa de manera correcta los procesos de generación del análisis?	4	3	5
<b>Promedio individual</b>		<b>4,66</b>	<b>3,66</b>	<b>4,83</b>
<b>Promedio total</b>		<b>4,38</b>		

**Tabla 17.** Evaluación del servicio de detección de organización de carteles

Participantes		P1, P2, P3 (Tabla 13)		
N°	Preguntas	Totalmente de acuerdo (5) De acuerdo (4) Indeciso (3) En desacuerdo (2) Totalmente en desacuerdo (1)		
		P1	P2	P3
1	¿Considera que la descripción del servicio mantiene una secuencia lógica y clara?	5	4	5
2	¿Considera que los diagramas que representan los procesos y subprocesos cumplen con las directrices de diseño sugeridas?	5	5	5
3	¿Considera que los diagramas que representan los procesos y subprocesos cumplen con las directrices de estilo de etiquetado sugeridas?	5	4	5
4	¿Considera que los diagramas que representan los procesos y subprocesos cumplen con las directrices de morfología sugeridas?	5	4	5
5	¿Considera que los diagramas que representan los procesos y subprocesos cumplen con las directrices de tamaño sugeridas?	5	4	5
6	¿Considera que el proceso de recolección de datos se ajusta a las necesidades del servicio?	4	5	4

7	¿Considera que los procesos relacionados en el servicio interactúan de manera correcta?	5	5	5
<b>Promedio individual</b>		<b>4,86</b>	<b>4,42</b>	<b>4,86</b>
<b>Promedio total</b>		<b>4,71</b>		

**Tabla 18.** Evaluación del servicio de alerta ciudadana

Participantes		P1, P2, P3 (Tabla 13)		
N°	Preguntas	Totalmente de acuerdo (5) De acuerdo (4) Indeciso (3) En desacuerdo (2) Totalmente en desacuerdo (1)		
		P1	P2	P3
1	¿Considera que la descripción del servicio mantiene una secuencia lógica y clara?	5	4	5
2	¿Considera que los diagramas que representan los procesos y subprocesos cumplen con las directrices de diseño sugeridas?	5	5	5
3	¿Considera que los diagramas que representan los procesos y subprocesos cumplen con las directrices de estilo de etiquetado sugeridas?	5	5	5
4	¿Considera que los diagramas que representan los procesos y subprocesos cumplen con las directrices de morfología sugeridas?	4	5	4
5	¿Considera que los diagramas que representan los procesos y subprocesos cumplen con las directrices de tamaño sugeridas?	4	5	5
6	¿Considera que los procesos relacionados en el servicio son los necesarios para llevar a cabo la gestión de las denuncias?	5	5	5
7	¿Considera que el servicio de alerta ciudadana se puede aplicar como herramienta para la recolección y gestión de denuncias en entidades estatales?	5	4	5
<b>Promedio individual</b>		<b>4,71</b>	<b>4,71</b>	<b>4,86</b>
<b>Promedio total</b>		<b>4,76</b>		

Con los resultados obtenidos en la evaluación final de cada uno de los servicios propuestos, se estableció que el modelo macro propuesto, cumple con los objetivos y directrices, en cuanto a diseño, concepto y estructura. Esta conclusión es preliminar ya que se hace necesario realizar su evaluación con un mayor número de participantes. Teniendo en cuenta los resultados mostrados en la **Tabla 15**, la **Tabla 16**, la **Tabla 17** y la **Tabla 18** se observa que el servicio de detección de similitudes contó con una calificación promedio de 4,55, que es excelente, teniendo en cuenta que en la escala usada, el 1 es totalmente en desacuerdo y 5 totalmente de acuerdo. Así mismo el servicio de detección de probable manipulación de ofertas contó con una calificación excelente de 4,38. El servicio de detección de carteles contó con una calificación excelente de 4,71 y finalmente el servicio de alerta ciudadana contó con una calificación excelente de 4,76. Los evaluadores además reportaron algunas observaciones para realizar mejoras a los modelos y estas fueron tenidos en cuenta para su presentación final en esta investigación.

## 5.2 Evaluación y experimentación de los servicios desarrollados

Teniendo en cuenta el modelo macro que se propuso y evaluó y los servicios que se implementaron (detección de similitud de propuestas técnicas y detección de manipulación de ofertas) a manera de prueba de concepto, se procedió a realizar la evaluación de estos dos servicios. Para esto y contemplando la complejidad de obtener información de propuestas que hayan sido tipificados en actos corruptos se siguió lo estipulado en el tercer objetivo específico que rige esta investigación, es decir, se procedió con la creación de datos sintéticos que sirvieron como fuente de información para realizar la mencionada evaluación.

### 5.2.1 Evaluación del servicio de detección de similitud

Para llevar a cabo la evaluación de este servicio se implementó una estrategia que consistió en poner a prueba diferentes escenarios que se pueden presentar al momento de comparar propuestas, como el copiado textual (sin agregar modificaciones), el copiado parcial (copiado solo de algunas frases) y el “copiado inteligente” (agregar modificaciones a frases con el objetivo de disimular la intención de copia). Algunos de los resultados arrojados en la evaluación se presentan en la **Figura 21**, la **Figura 22** y la **Figura 23**.

La **Figura 21** muestra los resultados de uno de los análisis de similitud realizados para el escenario de copiado textual, en este caso se cuenta con un documento que incluye frases copiadas textualmente desde diferentes fuentes de documentos. Los resultados arrojados por el servicio muestran que para todas las frases (9 en total) la probabilidad de similitud fue cercana al 95%, obteniendo una probabilidad de similitud del 99,64% para el documento en general.

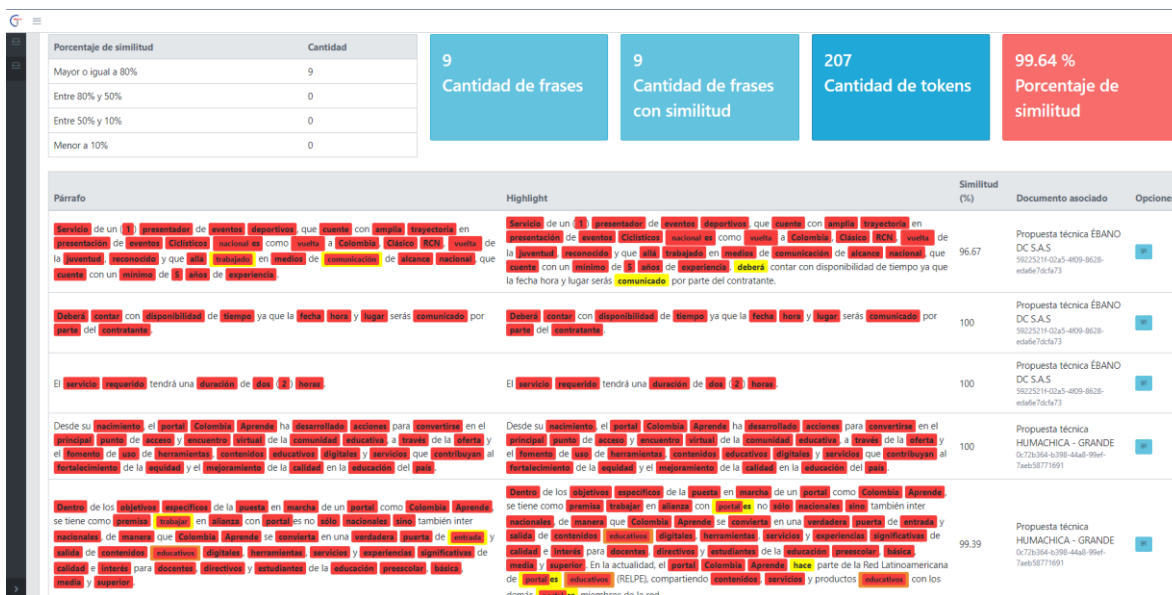
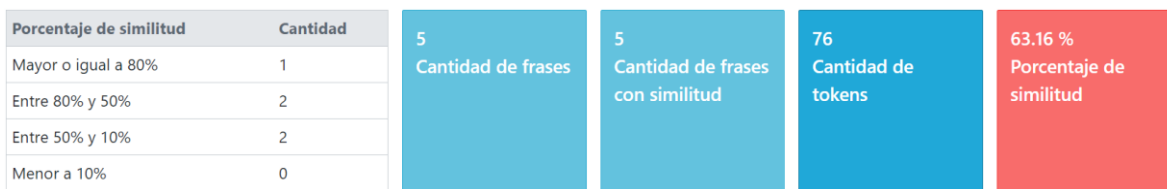


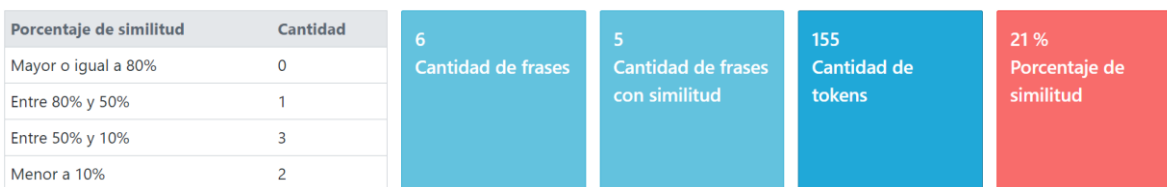
Figura 21. Resultados del análisis de similitud escenario copiado textual

La **Figura 22** corresponde a una de las pruebas realizadas para el escenario de copiado inteligente, en donde se cuenta con la intención de realizar plagio, pero esta vez, aplicando una estrategia de ocultamiento de este, haciendo uso de parafraseo, sinónimos o palabras similares, los resultados arrojados muestran que aún en este escenario el servicio logra detectar la similitud con una probabilidad considerable (63,16%).



**Figura 22.** Resultados del análisis de similitud escenario copiado inteligente

La **Figura 23** muestra una de las pruebas realizadas para el escenario en donde no se realiza plagio de manera intencional, para este caso, el servicio se comporta como se esperaría y arroja una probabilidad de similitud baja (21%).



**Figura 23.** Resultados del análisis de similitud escenario sin copia intencional

En total se realizaron pruebas con 6 convocatorias, cada una de ellas con 4 documentos de propuestas técnicas y con un total de 135 frases a evaluar. Con los resultados arrojados por la probabilidad de similitud de cada una de las frases evaluadas se procedió a la creación de la matriz de confusión y el cálculo de métricas de clásicas de evaluación en este tipo de soluciones.

El contenido de una frase es marcado como plagio, si su probabilidad de similitud es superior a un umbral dado. Para la correcta selección de dicho umbral, se procedió a realizar diferentes pruebas variando el valor de este. La **Figura 24** muestra el cálculo de algunas de las métricas más representativas del área evaluadas para diferentes valores del umbral. La experimentación realizada sugiere que un valor de 60% es el indicado para el conjunto de datos usados en la evaluación.

Los resultados finales de la experimentación se pueden apreciar en la **Figura 25**. Allí se logra apreciar que, si una propuesta técnica es marcada como plagio, existe un 82,7% de probabilidad de etiquetarla como tal. Mientras que, el servicio etiqueta al 20% de las que no son marcadas como plagio como partícipes de este posible delito (falsos positivos). Así mismo, la figura otorga información relevante sobre métricas calculadas a partir de la matriz de confusión, destacando, que el servicio obtiene una exactitud del 87% y una precisión del 96,23%. Los cuales son valores

aceptados por la comunidad de ciencia de datos como apropiados, no obstante, el tamaño de los datos de evaluación es muy pequeño y por esto se deben realizar futuras evaluaciones con conjuntos de mucho mayor tamaño.

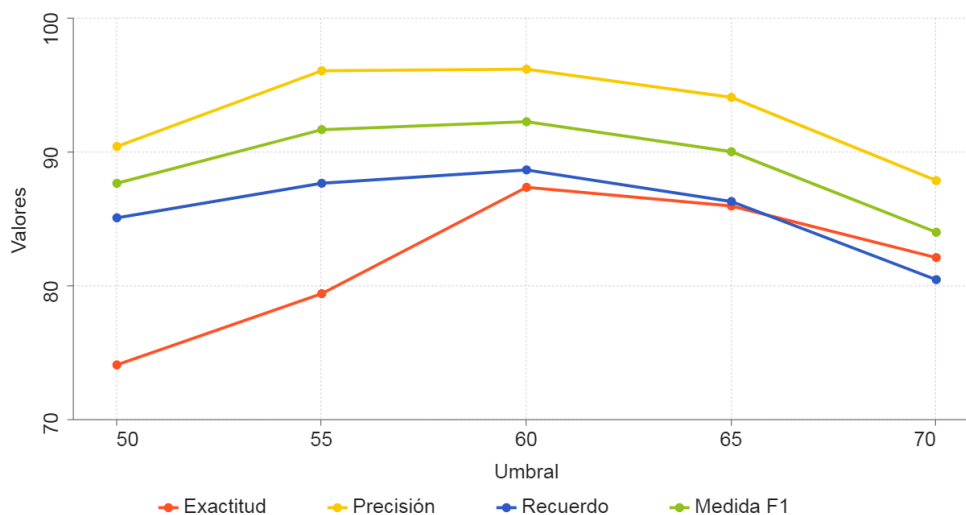


Figura 24. Gráfica de evaluación valores del umbral de detección de similitud

Matriz de confusión resultados de experimentación			Cálculo de métricas	
	Real verdadero	Real falso	Predicción general	
Predice Verdadero	102 (88,70%)	4 (20%)	106	Exactitud 87,40
Predice Falso	13 (11,30%)	16 (80%)	29	Precisión 96,23
Verdad general	115	20	<b>Total = 135</b>	Recuerdo 88,70
				Medida F1 92,30

Figura 25. Matriz de confusión y métricas sobre la experimentación del servicio de detección de similitud en propuestas técnicas

### 5.2.2 Evaluación del servicio de detección de manipulación de ofertas

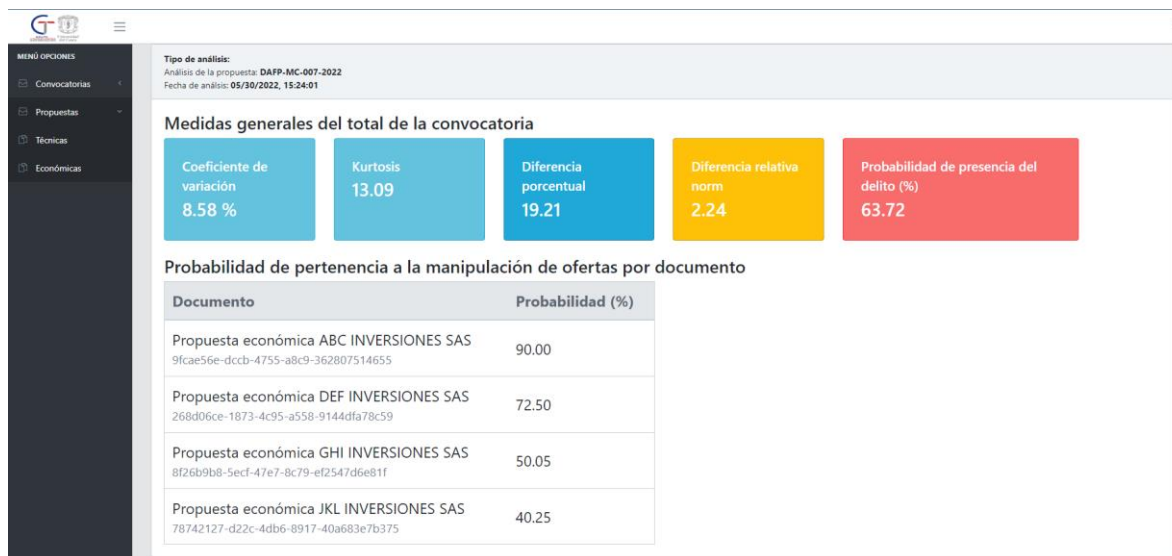
Para llevar a cabo la evaluación de este servicio se crearon diferentes escenarios de prueba que se pueden presentar en el delito de manipulación de ofertas dentro de una convocatoria. Algunos de los resultados se presentan a continuación en la Figura 20, la Figura 26 y la Figura 27.

La Figura 20 presentada anteriormente, en el apartado de **Desarrollo del servicio de detección de probable manipulación de ofertas**, muestra el escenario en donde 3 propuestas económicas hacen parte intencionalmente dentro del delito de manipulación de ofertas. El servicio en este caso arroja una probabilidad de



presencia del delito dentro de la convocatoria de un 65,47% y a las 3 propuestas las lista con una probabilidad superior al 75%.

La **Figura 26** a continuación muestra el escenario en donde 2 propuestas económicas son puestas intencionalmente dentro del delito de manipulación de ofertas, el servicio en este caso arroja una probabilidad de presencia del delito dentro de la convocatoria de un 63,72% y a los documentos partícipes del delito le otorga una probabilidad de participación superior del 70% mientras que a los 2 restantes su probabilidad fijada es igual o inferior al 50%.

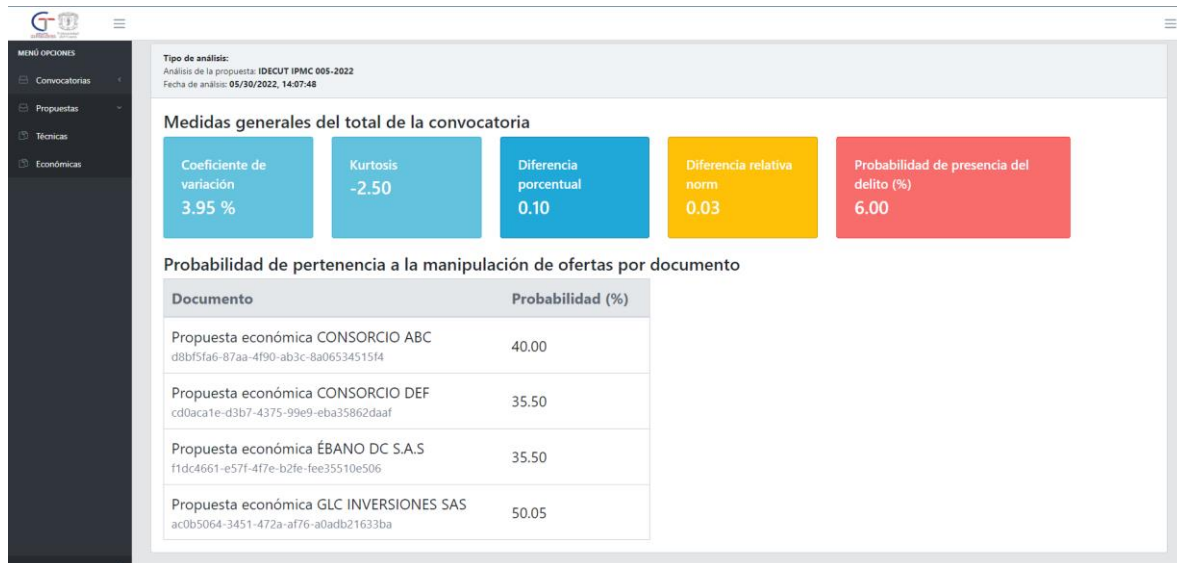


**Figura 26.** Resultados del análisis de manipulación de ofertas escenario dos documentos partícipes del delito

Finalmente, la **Figura 27** muestra una de las pruebas realizadas para el escenario en donde no se aplica de manera intencional el delito de manipulación de ofertas, el servicio se comporta como se esperaría y arroja una probabilidad de presencia del delito muy baja para la convocatoria (6%) mientras que, para cada propuesta, la probabilidad es igual o menor del 50%.

En total se realizaron pruebas con 8 convocatorias, cada una de ellas con 4 documentos de propuestas económicas, en donde se simularon doblemente 4 escenarios:

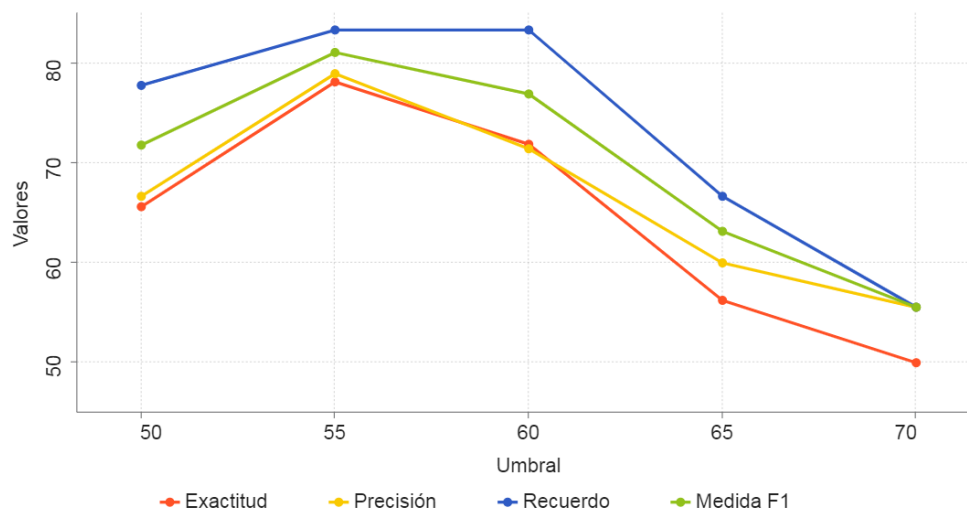
- Todos (4) los documentos hacen parte del delito.
- Tres (3) documentos hacen parte del delito.
- Dos (2) documentos hacen parte del delito.
- No hay presencia intencional del delito.



**Figura 27.** Resultados del análisis de manipulación de ofertas escenario de solo un documento partícipe del delito

Similar al proceso de evaluación del servicio de detección de similitudes, para esta evaluación se determina que la propuesta es marcada por el servicio como partícipe del delito de manipulación de ofertas si la probabilidad otorgada por el mismo supera un umbral fijado, y para la correcta selección de valor del umbral, se procedió a realizar el afinamiento de este, probando con diferentes valores. La

**Figura 28** muestra el cálculo de algunas de las métricas más representativas del área de ciencia de datos evaluadas para diferentes valores del umbral. El proceso de afinamiento del umbral sugiere que un valor de 55% es el indicado.



**Figura 28.** Gráfica de afinamiento del umbral de detección de manipulación de ofertas

Con los datos obtenidos en el afinamiento del umbral se procedió a la creación de la matriz de confusión y el cálculo de métricas clásicas del área resultantes de la evaluación, los cuales se pueden apreciar en la **Figura 29**.

Matriz de confusión resultados de experimentación				Cálculo de métricas	
	Real verdadero	Real falso	Predicción general	Métrica	Valor (%)
Predice Verdadero	15 (83,33%)	4 (28,60%)	19	Exactitud	78,13
Predice Falso	3 (16,67)	10 (71,40%)	13	Precisión	78,95
Verdad general	18	14	<b>Total = 32</b>	Recuerdo	83,33
				Medida F1	81,08

**Figura 29.** Matriz de confusión y métricas sobre la experimentación del servicio de detección de probable manipulación de ofertas

Se puede observar que, si una propuesta económica es marcada como partícipe del delito de manipulación de ofertas, existe un 83,3% de probabilidad de etiquetarla como tal. Mientras que, el servicio etiqueta al 28,6% de las que no son marcadas como partícipes del delito como sí (falsos positivos). Así mismo, la figura otorga información de métricas calculadas a partir de la matriz de confusión, destacando, que el servicio obtiene una exactitud de 78,13% y una precisión del 78,95%. Los cuales son buenos en el área de ciencia de datos, no obstante, se precisa realizar futuras evaluaciones con un mayor número de documentos, y que sean reales, no sintéticos.

Esta página ha sido intencionalmente dejada en blanco.

## CAPÍTULO 6

---

### 6. Conclusiones y trabajos futuros

#### 6.1 Conclusiones

Con respecto al primer objetivo específico, en esta investigación se realizó una revisión sistemática que buscó establecer los conceptos fundamentales del estado de arte relacionados con soluciones que apliquen estrategias anticorrupción y con base en esto, analizar la viabilidad de la creación de un modelo de detección e identificación de corrupción en el proceso de contratación pública. Los resultados obtenidos con la revisión sistemática de la literatura permitieron concluir que el tema abordado en esta investigación es un problema que en los últimos años es cada vez más interesante para la comunidad académica y científica, y que expertos en el área recomiendan la implementación de diversas estrategias para atacar el problema, como el uso de técnicas estadísticas y la combinación de estas técnicas con tecnologías de la información. Con la RSL, además, se lograron obtener pautas significativas para la posterior construcción del modelo desarrollado en esta tesis.

En relación con el segundo objetivo específico se realizó un modelo macro que consta de, 1) servicios especializados diseñados para abordar una situación específica donde hay probabilidad de corrupción, en donde se plantearon tres servicios dedicados a tres delitos: la detección de similitud entre propuestas económicas, la detección de la manipulación de ofertas y la detección de carteles, 2) servicios transversales que soportan la transformación del modelo en una herramienta software, donde se plantea la implementación de los servicios de administración, autenticación y autorización de usuarios y la administración de los servicios, 3) servicios adicionales, que pueden, sin necesidad de identificar un delito específico, ayudar a la identificación de situaciones anómalas en un proceso licitatorio, planteando para este componente el servicio de alerta ciudadana, 4) las relaciones explícitas entre los servicios y 5) la salida global del modelo. La representación visual se efectuó de acuerdo con los estándares de modelado en nomenclatura UML. El modelo en general busca, de manera conjunta con los servicios planteados, detectar diferentes probables delitos o anomalías relacionadas con corrupción en los procesos de contratación pública en Colombia.

El modelo propuesto en esta investigación fue sujeto a realimentación a través de un grupo focal con tres (3) expertos y posteriormente fue evaluado con otros tres (3) expertos. Los expertos seleccionados tienen conocimientos y experiencia en estrategias anticorrupción o en tecnologías de la información. Esta evaluación permitió en primer lugar, identificar aspectos por mejorar en la presentación de los procesos que involucran a los servicios y al modelo propuesto, para una mayor

compresión, y a confirmar que el camino que se estaba siguiendo en la propuesta era el correcto. Y en un segundo lugar, la evaluación formal del modelo, con base en las directrices definidas por Montes de Oca, permitió definir el correcto nivel de representación visual y complejidad de los procesos descritos por los diagramas que soportan el modelo, y a que cada proceso cumpliera con los objetivos para los que fue diseñado. Los resultados de la evaluación fueron positivos, obteniendo, de acuerdo con la escala calificativa de Likert, resultados de **Totalmente de acuerdo** y **De acuerdo**, con un valor superior a 4,3 sobre 5 en todos los servicios.

Teniendo en cuenta lo mencionado previamente se concluye que el modelo propuesto en esta investigación cumple preliminarmente (se hace necesario realizar su evaluación con un mayor número de participantes) con los objetivos y directrices, en cuanto a diseño, concepto y estructura. Así mismo, se logró establecer, gracias a las revisiones y evaluaciones hechas, que los datos de entrada, los procesos internos y las salidas establecidas para cada uno de los servicios son los requeridos para cumplir con el objetivo que tienen cada uno de los servicios.

En relación con el tercer objetivo específico, se creó una herramienta software, que sirvió como prueba de concepto del modelo propuesto. La herramienta se basa en los componentes transversales del modelo propuesto y dos servicios especializados en la detección de posibles delitos de corrupción. Los servicios especializados fueron: el **servicio de detección de similitud en propuestas técnicas** y el **servicio de detección de probable manipulación de ofertas**. Lo anterior se logró desarrollando el backend de los servicios con librerías de Python y el respectivo frontend y la conexión global con el Framework Angular.

Debido a la complejidad de obtener información de documentos que hayan sido tipificados en actos corruptos, en esta investigación y con el objetivo de llevar a cabo la evaluación de los servicios implementados se creó un conjunto de datos sintéticos. La calidad de los servicios fue evaluada mediante métricas clásicas del área, precisión, recuerdo, medida F1 y la exactitud, obteniendo valores relativamente altos.

En relación con el tercer objetivo específico se concluye que los dos servicios especializados desarrollados permiten demostrar que la implementación de estos es viable, así mismo, gracias a la evaluación realizada sobre estos, se puede concluir (preliminarmente) que los dos servicios permiten la identificación de presuntas situaciones de corrupción asociados a los procesos de contratación pública con un alto grado de precisión y exactitud.

En relación con la hipótesis de investigación (H1) planteada al inicio de la investigación se logra dar una respuesta afirmativa a esta, al proponer un modelo que incluye tres (3) servicios especializados y un (1) servicio complementario, dedicados a la tarea de la identificación temprana de presuntas situaciones de corrupción en procesos de contratación estatal en Colombia, junto con la descripción detallada de los procesos que estos deben llevar a cabo para cumplir con sus objetivos. La respuesta es afirmativa debido a que los servicios planteados

en el modelo buscan atacar los delitos de corrupción mayormente abordados en investigaciones previas y los resultados de su evaluación fueron prometedores, sin embargo, es válido mencionar que, con el fin de enriquecer aún más el modelo propuesto, se deberían plantear nuevos servicios especializados que busquen atacar posibles delitos de corrupción que no son abordados por los servicios propuestos en esta investigación, esto es sugerido como trabajo futuro en esta tesis.

## 6.2 Trabajos futuros

Como trabajo futuro, se considera importante la integración de nuevos servicios específicos dedicados a la detección de nuevos delitos. Así como la implementación del **servicio especializado de detección del probable delito de organización de carteles y el análisis del beneficiario final** y del **servicio de alerta ciudadana sobre irregularidades en procesos de convocatorias de contratación pública**. Esto con el objetivo de enriquecer tanto el modelo propuesto en esta investigación, como la herramienta software que lo soporta.

Finalmente, el grupo de investigación espera poder desplegar la herramienta software desarrollada en una entidad estatal colombiana, integrándola en sus procesos de contratación, para permitir brindar una ayuda en la tarea de la prevención de la corrupción en los procesos de contratación estatal.

## CAPÍTULO 7

---

### 7. Bibliografía

- [1] T. por Colombia, “Índice de Percepción de la Corrupción 2021-Transparencia por Colombia,” 2021.
- [2] A. del Castillo, C. A. Rodríguez, L. H. González, and Z. M. Prada, *La corrupción en la contratación estatal: ¿entre la etiqueta y la acción social?*, no. 86. Auditoría Superior de la República, 2017. doi: <https://doi.org/10.1162/003465303772815871>.
- [3] L. Froeb, D. Sibley, M. Doane, and B. P. Pinto, “Screening for collusion as a problem of inference,” in *The Oxford Handbook of International Antitrust Economics, Volume 2*, 2014.
- [4] O. de coopération et de développement économiques, *Reforming Public Procurement: Progress in Implementing the 2015 OECD Recommendation*. OECD Publishing, 2019.
- [5] N. Ehlermann-Cache, *Bribery in public procurement: methods, actors and counter-measures*. OECD, 2007.
- [6] D. Peltier-Rivest, “The prevention and detection of corruption in pharmaceutical companies,” *Pharmaceuticals Policy and Law*, vol. 19, no. 1–2, pp. 17–31, 2017, doi: <https://dx.doi.org/10.3233/PPL-170451>.
- [7] D. Peltier-Rivest, “A model for preventing corruption,” *Journal of Financial Crime*, vol. 25, no. 2, pp. 545–561, 2018, doi: <https://doi.org/10.1108/JFC-11-2014-0048>.
- [8] R. H. Porter and J. D. Zona, “Detection of bid rigging in procurement auctions,” *Journal of political economy*, vol. 101, no. 3, pp. 518–538, 1993, doi: <https://doi.org/10.1086/261885>.
- [9] I. Morozov and E. A. Podkolzina, “Collusion detection in procurement auctions,” *Higher School of Economics Research Paper No. WP BPR*, vol. 25, 2013, doi: <http://dx.doi.org/10.2139/ssrn.2221809>.
- [10] D. Imhof, Y. Karagök, and S. Rutz, “Screening for Bid Rigging—Does It Work?,” *Journal of Competition Law & Economics*, vol. 14, no. 2, pp. 235–261, 2018, doi: <https://doi.org/10.1093/joclec/nhy006>.
- [11] R. Chotibhongs and D. Arditi, “Detection of collusive behavior,” *Journal of Construction Engineering and Management*, vol. 138, no. 11, pp. 1251–1258, 2012, doi: [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0000542](https://doi.org/10.1061/(ASCE)CO.1943-7862.0000542).
- [12] P. Bajari and L. Ye, “Deciding between competition and collusion,” *Review of Economics and Statistics*, vol. 85, no. 4, pp. 971–989, 2003.
- [13] M. Shan, A. P. C. Chan, Y. Le, B. Xia, and Y. Hu, “Measuring corruption in public construction projects in China,” *Journal of Professional Issues in Engineering Education and Practice*, vol. 141, no. 4, p. 5015001, 2015, doi: [https://doi.org/10.1061/\(ASCE\)EI.1943-5541.0000241](https://doi.org/10.1061/(ASCE)EI.1943-5541.0000241).



- [14] K. A. Belokrylov and L. P. Runova, "Localization of Subjective Assessment Risks in the Public Procurement System Based on Fuzzy Logic," *Journal of Advanced Research in Law and Economics*, vol. 8, no. 2 (24), pp. 411–417, 2017, doi: [https://dx.doi.org/10.14505//jarle.v8.2\(24\).09](https://dx.doi.org/10.14505//jarle.v8.2(24).09).
- [15] E. K. Owusu, A. P. C. Chan, and E. Ameyaw, "Toward a cleaner project procurement: Evaluation of construction projects' vulnerability to corruption in developing countries," *J Clean Prod*, vol. 216, pp. 394–407, 2019, doi: <https://doi.org/10.1016/j.jclepro.2019.01.124>.
- [16] S. T. Parente, B. Schulte, A. Jost, T. Sullivan, and A. Klindworth, "Assessment of predictive modeling for identifying fraud within the Medicare program," *Health Manag Policy Innov*, vol. 1, no. 2, pp. 8–37, 2012, [Online]. Available: <https://hmpi.org/wp-content/uploads/2017/02/HMPI-Parente-Fortel-Anyalytics-LLC-Fraud-PreventManu.pdf>
- [17] M. Fazekas, I. J. Tóth, and L. P. King, "An objective corruption risk index using public procurement data," *European Journal on Criminal Policy and Research*, vol. 22, no. 3, pp. 369–397, 2016, doi: <https://doi.org/10.1007/s10610-016-9308-z>.
- [18] M. Huber and D. Imhof, "Machine learning with screens for detecting bid-rigging cartels," *International Journal of Industrial Organization*, vol. 65, pp. 277–301, 2019, doi: <https://doi.org/10.1016/j.ijindorg.2019.04.002>.
- [19] M. T. Islam and M. A. Yousuf, "Development of a Corruption Detection Algorithm using K-means Clustering," in *2018 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*, 2018, pp. 1–4. doi: <https://doi.org/10.1109/ICAEEE.2018.8642985>.
- [20] C. G. Ralha and C. V. S. Silva, "A multi-agent data mining system for cartel detection in Brazilian government procurement," *Expert Systems with Applications*, vol. 39, no. 14, pp. 11642–11656, 2012, doi: <https://doi.org/10.1016/j.eswa.2012.04.037>.
- [21] F. J. López-Iturriaga and I. P. Sanz, "Predicting public corruption with neural networks: An analysis of spanish provinces," *Social Indicators Research*, vol. 140, no. 3, pp. 975–998, 2018, doi: <https://doi.org/10.1007/s11205-017-1802-2>.
- [22] J. Wachs and J. Kertész, "A network approach to cartel detection in public auction markets," *arXiv preprint arXiv:1906.08667*, 2019, doi: <https://doi.org/10.1038/s41598-019-47198-1>.
- [23] J. A. Dorofeeva and M. A. Tokmakov, "The Use of Identical IP Address by Tender Participants as an Indication of Cartel Scheme," in *Digital Transformation of the Economy: Challenges, Trends and New Opportunities*, Springer, 2020, pp. 244–252. doi: [https://doi.org/10.1007/978-3-030-11367-4\\_23](https://doi.org/10.1007/978-3-030-11367-4_23).
- [24] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering--a systematic literature review," *Inf Softw Technol*, vol. 51, no. 1, pp. 7–15, 2009, doi: <https://doi.org/10.1016/j.infsof.2008.09.009>.
- [25] K. S. Pratt, "Design Patterns for Research Methods: Iterative Field Research," 2009.

- [26] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," 2007.
- [27] J. Bailey, D. Budgen, M. Turner, B. Kitchenham, P. Brereton, and S. Linkman, "Evidence relating to Object-Oriented software design: A survey," in *First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007)*, 2007, pp. 482–484.
- [28] P. Bajari, "Comparing competition and collusion: a numerical approach," *Economic Theory*, vol. 18, no. 1, pp. 187–205, 2001, doi: <https://doi.org/10.1007/PL00004128>.
- [29] M. Oraee, M. R. Hosseini, D. J. Edwards, H. Li, E. Papadonikolaki, and D. Cao, "Collaboration barriers in BIM-based construction networks: A conceptual model," *International Journal of Project Management*, vol. 37, no. 6, pp. 839–854, Aug. 2019, doi: 10.1016/J.IJPROMAN.2019.05.004.
- [30] E. K. Owusu, A. P. C. Chan, and M. Shan, "Causal Factors of Corruption in Construction Project Management: An Overview," *Science and Engineering Ethics*, vol. 25, no. 1, pp. 1–31, Feb. 2019, doi: 10.1007/S11948-017-0002-4/TABLES/5.
- [31] ImranNaveed, "Electronic media, creativity and plagiarism," *ACM SIGCAS Computers and Society*, vol. 40, no. 4, pp. 25–44, Dec. 2010, doi: 10.1145/1929609.1929613.
- [32] F. Sánchez-Vega, E. Villatoro-Tello, M. Montes-Y-Gómez, L. Villaseñor-Pineda, and P. Rosso, "Determining and characterizing the reused text for plagiarism detection," *Expert Systems with Applications*, vol. 40, no. 5, pp. 1804–1813, Apr. 2013, doi: 10.1016/J.ESWA.2012.09.021.
- [33] B. Stein, S. M. Zu Eissen, and M. Potthast, "Strategies for retrieving plagiarized documents," *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'07*, pp. 825–826, 2007, doi: 10.1145/1277741.1277928.
- [34] K. Vani and D. Gupta, "Detection of idea plagiarism using syntax–Semantic concept extractions with genetic algorithm," *Expert Systems with Applications*, vol. 73, pp. 11–26, May 2017, doi: 10.1016/J.ESWA.2016.12.022.
- [35] FoltýnekTomáš, MeuschkeNorman, and GippBela, "Academic Plagiarism Detection," *ACM Computing Surveys (CSUR)*, vol. 52, no. 6, Oct. 2019, doi: 10.1145/3345317.
- [36] M. Y. M. Chong, "A study on plagiarism detection and plagiarism direction identification using natural language processing techniques," 2013, Accessed: Oct. 18, 2021. [Online]. Available: <https://wlv.openrepository.com/handle/2436/298219>
- [37] M. Mohebbi and A. Talebpour, "Texts semantic similarity detection based graph approach.," *Int. Arab J. Inf. Technol.*, vol. 13, no. 2, pp. 246–251, 2016.
- [38] T. A. E. Eisa, N. Salim, and S. Alzahrani, "Existing plagiarism detection techniques: A systematic mapping of the scholarly literature," *Online Information Review*, vol. 39, no. 3, pp. 383–400, Jun. 2015, doi: 10.1108/OIR-12-2014-0315.
- [39] N. Meuschke, V. Stange, M. Schubotz, M. Kramer, and B. Gipp, "Improving academic plagiarism detection for STEM documents by analyzing mathematical content and citations," *Proceedings of the ACM/IEEE Joint*

- Conference on Digital Libraries*, vol. 2019-June, pp. 120–129, Jun. 2019, doi: 10.1109/JCDL.2019.00026.
- [40] W. B. Priambodo and G. Guarddin, “Federated document search using hybrid peer-to-peer network for plagiarism detection,” *2016 International Conference on Advanced Computer Science and Information Systems, ICACIS 2016*, pp. 61–66, Mar. 2017, doi: 10.1109/ICACIS.2016.7872748.
- [41] E. Băutu and A. Băutu, “PlagZap: A Textual Plagiarism Detection System for Student Assignments Built with Open-Source Software,” *Advances in Intelligent Systems and Computing*, vol. 771, pp. 500–508, Jun. 2018, doi: 10.1007/978-3-319-94120-2\_48.
- [42] N. Meuschke, V. Stange, M. Schubotz, and B. Gipp, “Hyplag: A hybrid approach to academic plagiarism detection,” *41st International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2018*, pp. 1321–1324, Jun. 2018, doi: 10.1145/3209978.3210177.
- [43] B. Gipp, N. Meuschke, C. Breitingner, J. Pitman, and A. Nürnberger, “Web-based Demonstration of Semantic Similarity Detection Using Citation Pattern Visualization for a Cross Language Plagiarism Case.,” in *ICEIS (2)*, 2014, pp. 677–683.
- [44] N. Meuschke, N. Siebeck, M. Schubotz, and B. Gipp, “Analyzing semantic concept patterns to detect academic plagiarism,” *ACM International Conference Proceeding Series*, pp. 46–53, Dec. 2017, doi: 10.1145/3127526.3127535.
- [45] A. S. Rodríguez, “El plagio y su impacto a nivel académico y profesional,” *e-Ciencias de la Información*, vol. 2, no. 1, pp. 1–13, Aug. 2012, doi: 10.15517/ECI.V2I1.1213.
- [46] B. Stein, S. M. Zu Eissen, and M. Potthast, “Strategies for retrieving plagiarized documents,” *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR’07*, pp. 825–826, 2007, doi: 10.1145/1277741.1277928.
- [47] G. E. Pibiri and R. Venturini, “Techniques for Inverted Index Compression,” *ACM Computing Surveys (CSUR)*, vol. 53, no. 6, Dec. 2020, doi: 10.1145/3415148.
- [48] C. E. Ossa Bocanegra, “Tratamiento de la colusión en la contratación pública: una visión del caso colombiano,” *Revista de Derecho*, no. 42, pp. 233–263, 2014.
- [49] A. Oke, C. Aigbavboa, and Z. Mangena, “Prevention of collusion for innovative construction,” *Procedia Eng*, vol. 196, pp. 491–497, 2017.
- [50] A. Foremny, “Review of collusion and bid rigging detection methods in the construction industry,” in *Creative Construction Conference 2018*, 2018, pp. 946–953.
- [51] H. Wallimann, D. Imhof, and M. Huber, “A Machine Learning Approach for Flagging Incomplete Bid-rigging Cartels,” Apr. 2020, Accessed: Oct. 05, 2021. [Online]. Available: <https://arxiv.org/abs/2004.05629v1>
- [52] I. Safder, S. U. Hassan, A. Visvizi, T. Noraset, R. Nawaz, and S. Tuarob, “Deep Learning-based Extraction of Algorithmic Metadata in Full-Text Scholarly Documents,” *Information Processing & Management*, vol. 57, no. 6, p. 102269, Nov. 2020, doi: 10.1016/J.IPM.2020.102269.

- [53] M. Hu *et al.*, “CNN-IETS: A CNN-based probabilistic approach for information extraction by text segmentation,” *International Conference on Information and Knowledge Management, Proceedings*, vol. Part F131841, pp. 1159–1168, Nov. 2017, doi: 10.1145/3132847.3132962.
- [54] D. Imhof, “Empirical Methods for Detecting Bid-rigging Cartels,” *Tesis Doctoral. Bourgogne Franche-Comté*, Oct. 2018.
- [55] G. Erkan and D. R. Radev, “LexRank: Graph-based Lexical Centrality as Saliency in Text Summarization,” *Journal of Artificial Intelligence Research*, vol. 22, pp. 457–479, Dec. 2004, doi: 10.1613/JAIR.1523.
- [56] A. P. Lleras, “La lucha contra los carteles empresariales en la contratación estatal en Colombia\*,” <https://doi.org/10.29263/lar03.2019.06>, no. 3, pp. 117–134, Aug. 2019, doi: 10.29263/LAR03.2019.06.
- [57] S. Industria y Comercio, “Combatir la colusión en las licitaciones,” 2013. [https://www.sic.gov.co/recursos\\_user/documentos/articulos/2010/Guia\\_Contratacion.pdf](https://www.sic.gov.co/recursos_user/documentos/articulos/2010/Guia_Contratacion.pdf) (accessed Feb. 07, 2022).
- [58] T. O. OECD, “Collusion and Corruption in Public Procurement.” París: OECD, 2010.
- [59] P. Bajari and G. Summers, “Detecting Collusion in Procurement Auctions,” *Antitrust Law Journal*, vol. 70, 2002, Accessed: Nov. 09, 2021. [Online]. Available: <https://heinonline.org/HOL/Page?handle=hein.journals/antil70&id=153&div=&collection=>
- [60] K. Brisset and L. Thomas, “Leniency Program: A New Tool in Competition Policy to Deter Cartel Activity in Procurement Auctions,” *European Journal of Law and Economics* 2004 17:1, vol. 17, no. 1, pp. 5–19, Jan. 2004, doi: 10.1023/A:1026329724892.
- [61] R. A. Baldomir, G. C. G. van Erven, and C. G. Ralha, “Brazilian Government Procurements: an Approach to Find Fraud Traces in Companies Relationships,” *Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*, pp. 752–762, Oct. 2018, doi: 10.5753/ENIAC.2018.4464.
- [62] K. Rabuzin and N. Modrusan, “Prediction of Public Procurement Corruption Indices using Machine Learning Methods.,” in *KMIS*, 2019, pp. 333–340.
- [63] D. Imhof *et al.*, “Detecting bid-rigging coalitions in different countries and auction formats,” *International Review of Law and Economics*, vol. 68, May 2021, doi: 10.1016/j.irl.2021.106016.
- [64] M. Huber and D. Imhof, “Deep learning for detecting bid rigging: Flagging cartel participants based on convolutional neural networks,” Apr. 2021, Accessed: Nov. 10, 2021. [Online]. Available: <https://arxiv.org/abs/2104.11142v1>
- [65] P. M. Gilmour, “Lifting the veil on beneficial ownership: Challenges of implementing the UK’s registers of beneficial owners,” *Journal of Money Laundering Control*, vol. 23, no. 4, pp. 717–734, Dec. 2020, doi: 10.1108/JMLC-02-2020-0014/FULL/XML.
- [66] H. A. Hernández Useche, “Mejores prácticas para identificación del beneficiario final,” 2017, Accessed: Feb. 07, 2022. [Online]. Available:

- <https://isbn.cloud/9789589040713/mejores-practicas-para-la-identificacion-del-beneficiario-final/>
- [67] M. Riccardi and V. Agostino, "Identifying Beneficial Owners for Fighting against Money Laundering: The Result of EU Project Bownet," *European Police Science and Research Bulletin*, vol. 8, 2013, Accessed: Nov. 17, 2021. [Online]. Available: <https://heinonline.org/HOL/Page?handle=hein.journals/elerb8&id=14&div=&collection=>
- [68] D. Imhof, "DETECTING BID-RIGGING CARTELS WITH DESCRIPTIVE STATISTICS," *Journal of Competition Law & Economics*, vol. 15, no. 4, pp. 427–467, Dec. 2019, doi: 10.1093/JOCLEC/NHZ019.
- [69] R. M. Abrantes-Metz, "Roundtable on Ex Officio Cartel Investigations and the Use of Screens to Detect Cartels," *SSRN Electronic Journal*, Oct. 2013, doi: 10.2139/SSRN.2343465.
- [70] E. E. Ameyaw, E. Pärn, A. P. C. Chan, D.-G. Owusu-Manu, D. J. Edwards, and A. Darko, "Corrupt Practices in the Construction Industry: Survey of Ghanaian Experience," *Journal of Management in Engineering*, vol. 33, no. 6, p. 05017006, Nov. 2017, doi: 10.1061/(ASCE)ME.1943-5479.0000555.
- [71] A. P. C. Chan and E. K. Owusu, "Corruption Forms in the Construction Industry: Literature Review," *Journal of Construction Engineering and Management*, vol. 143, no. 8, p. 04017057, Jun. 2017, doi: 10.1061/(ASCE)CO.1943-7862.0001353.
- [72] L. Farca, "Conflict of Interests Under Romanian Public Procurement Rules. Pros and Cons of the Prevent System," *SSRN Electronic Journal*, Jul. 2018, doi: 10.2139/SSRN.3246863.
- [73] S. E. Trapnell, "Users' guide to measuring corruption and anti-corruption. United Nations Development Programme." 2015.
- [74] E. Kingsford Owusu and A. P. C. Chan, "Barriers Affecting Effective Application of Anticorruption Measures in Infrastructure Projects: Disparities between Developed and Developing Countries," *Journal of Management in Engineering*, vol. 35, no. 1, p. 04018056, Jan. 2019, doi: 10.1061/(ASCE)ME.1943-5479.0000667.
- [75] M. Thierry and T. K. Priyambodo, "SMS and Web-Based e-Government Model Case Study: Citizens Complaints Management System at District of Gihosha –Burundi," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 11, no. 1, pp. 67–76, Jan. 2017, doi: 10.22146/IJCCS.17167.
- [76] K. H. Mohammed, A. Hassan, and D. Yusuf Mohammed, "Identity and Access Management System: a Web-Based Approach for an Enterprise," *Traektorîa Nauki = Path of Science. 2018*, vol. 4, no. 11, 2018, doi: 10.22178/pos.40-1.
- [77] J. P. Cruz, Y. Kaji, and N. Yanai, "RBAC-SC: Role-based access control using smart contract," *IEEE Access*, vol. 6, pp. 12240–12251, Mar. 2018, doi: 10.1109/ACCESS.2018.2812844.
- [78] K. Sharan, "Model-View-Controller Pattern," *Learn JavaFX 8*, pp. 419–434, 2015, doi: 10.1007/978-1-4842-1142-7\_11.
- [79] M. Attaran and J. Woods, "Cloud computing technology: improving small business performance using the Internet,"

- <https://doi.org/10.1080/08276331.2018.1466850>, vol. 31, no. 6, pp. 495–519, Nov. 2018, doi: 10.1080/08276331.2018.1466850.
- [80] I. Weber, Q. Lu, A. B. Tran, A. Deshmukh, M. Gorski, and M. Strazds, “A platform architecture for multi-tenant blockchain-based systems,” *Proceedings - 2019 IEEE International Conference on Software Architecture, ICSA 2019*, pp. 101–110, Apr. 2019, doi: 10.1109/ICSA.2019.00019.
- [81] M. Waseem, P. Liang, and M. Shahin, “A Systematic Mapping Study on Microservices Architecture in DevOps,” *Journal of Systems and Software*, vol. 170, p. 110798, Dec. 2020, doi: 10.1016/J.JSS.2020.110798.
- [82] S. Peltonen, L. Mezzalana, and D. Taibi, “Motivations, benefits, and issues for adopting Micro-Frontends: A Multivocal Literature Review,” *Information and Software Technology*, vol. 136, p. 106571, Aug. 2021, doi: 10.1016/J.INFSOF.2021.106571.
- [83] V. Gkantouna, V. Papaioannou, G. Tzimas, and Z. Sabic, “A Semantic Approach for Domain-Specific Design Patterns Recommendations in CMS-based Web Development,” *INISTA 2020 - 2020 International Conference on INnovations in Intelligent SysTems and Applications, Proceedings*, Aug. 2020, doi: 10.1109/INISTA49547.2020.9194622.
- [84] A. Téllez-Valero, M. Montes-y-Gómez, and L. Villaseñor-Pineda, “A Machine Learning Approach to Information Extraction,” *Lecture Notes in Computer Science*, vol. 3406, pp. 539–547, 2005, doi: 10.1007/978-3-540-30586-6\_58.
- [85] M. MENDOZA-MORENO, C. GONZÁLEZ-SERRANO, and F. J. PINO, “FOCUS GROUP COMO PROCESO EN INGENIERÍA DE SOFTWARE: UNA EXPERIENCIA DESDE LA PRÁCTICA,” *DYNA*, vol. 80, no. 181, pp. 51–60, Sep. 2013, Accessed: Apr. 22, 2022. [Online]. Available: <https://revistas.unal.edu.co/index.php/dyna/article/view/36969>
- [86] Moreno Montes de Oca, “Patrón y clasificación taxonómica para directrices prácticas en modelos de procesos de negocio,” Universidad Central “Marta Abreu” de Las Villas., 2015.