

Escuela Superior Politécnica del Litoral

Facultad de Ingeniería en Electricidad y Computación

Caracterización de proveedores de bienes y servicios para un sistema de compras
públicas

Proyecto Integrador

Previo la obtención del Título de:

Ingeniero en Computación

Presentado por:

Natalia Romina Ramírez Yépez

Emanuel Alejandro Parra Bagua

Guayaquil - Ecuador

Año: 2023

Dedicatoria

A mi querida madre. Aunque ya no estás físicamente conmigo, siento tu presencia en cada desafío superado y en cada logro alcanzado. Esta tesis es un reflejo del amor, la sabiduría y la tenacidad que me inculcaste. A través de estas páginas, tu legado vive y brilla.

Te dedico este trabajo, con la esperanza de que, desde donde quiera que estés, puedas sentir el profundo amor y gratitud que llevo en mi corazón. Tu memoria ha sido mi guía y mi fortaleza.

Emanuel Alejandro Parra Bagua

Dedicatoria

Para mi abuela, que para mí y para muchos fuiste como una segunda madre. Estuviste conmigo cuando empecé este capítulo tan importante y, aunque te fuiste antes de que pudiese terminarlo, el gran ejemplo que me dejaste me dio la motivación necesaria para llegar a este momento.

Te dedico este trabajo como agradecimiento a todo tu esfuerzo y amor. Fui tu primera nieta en nacer, y también la primera en obtener un título. Sé que, donde quiera que estés, estás orgullosa de mí.

Natalia Romina Ramírez Yépez

Agradecimientos

Mi más sincero agradecimiento a Dios, quien es el encargado de guiar nuestros caminos. A mi familia, por su constante apoyo, ánimo y comprensión. A nuestro tutor, quien puso a nuestra disposición su tiempo, guía, paciencia y dedicación en la finalización de este viaje académico. A Andrea, mi compañera de vida, tu amor, apoyo y fe en mí han sido inquebrantables.

A todos ustedes, mi más profundo y sincero agradecimiento.

Emanuel Alejandro Parra Bagua

Agradecimientos

Quiero expresar mi profundo agradecimiento a mi familia, a mis padres, por su apoyo incondicional durante todos estos años. A nuestro tutor, quien fue nuestra constante guía en el proceso de esta investigación. A mis amigos, quienes estuvieron ahí siempre para apoyarme emocionalmente. A Sebastian, por comprender mis momentos de estrés, por hacer que me supere a mí misma y por alentarme a seguir adelante. A todos ustedes, gracias.

Natalia Romina Ramírez Yépez

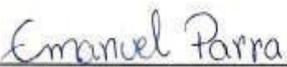
Declaración Expresa

Yo/Nosotros *Emanuel Alejandro Parra Bagua* y *Natalia Romina Ramírez Yépez* acuerdo/acordamos y reconozco/reconocemos que la titularidad total y exclusiva sobre los derechos patrimoniales de patente de invención, modelo de utilidad, diseño industrial, información no divulgada y cualquier otro derecho o tipo de Propiedad Intelectual que corresponda o pueda corresponder respecto de cualquier investigación, desarrollo tecnológico o invención realizada durante el desarrollo de su trabajo de titulación, incluyendo cualquier derecho de participación de beneficios o de valor sobre titularidad de derechos, pertenecerán de forma total, perpetua, exclusiva e indivisible a LA ESPOL, sin limitación de ningún tipo. Se deja además expresa constancia de que lo aquí establecido constituye un “previo acuerdo”, así como de ser posible bajo la normativa vigente de transferencia o cesión a favor de la ESPOL de todo derecho o porcentaje de titularidad que pueda existir.

Sin perjuicio de lo anterior los alumnos firmantes de la presente declaración reciben en este acto una licencia de uso gratuita e intransferible de plazo indefinido para el uso no comercial de cualquier investigación, desarrollo tecnológico o invención realizada durante el desarrollo de su trabajo de titulación, sin perjuicio de lo cual deberán contar con una autorización previa expresa de la ESPOL para difundir públicamente el contenido de la investigación, desarrollo tecnológico o invención.

Así también autorizamos expresamente a que la ESPOL realice la comunicación pública de la obra o invento, por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual.

Guayaquil, 2 de septiembre del 2023.


Emanuel Alejandro

Parra Bagua



Natalia Romina Ramírez

Yépez

Evaluadores

Lucía Marisol Villacrés Falconi

Profesor de Materia

Eduardo Segundo Cruz Ramírez

Tutor de proyecto

Resumen

El sistema de contratación pública en Ecuador, SERCOP, ha enfrentado numerosos desafíos, incluidos la corrupción y la ineficiencia, lo que ha llevado a la pérdida de una cantidad significativa de dinero público. A pesar de los esfuerzos por mejorar la transparencia y la eficiencia, aún existen irregularidades e inconsistencias en los procesos de contratación. Bajo esta premisa se propone desarrollar un modelo de caracterización de proveedores de bienes y servicios del SERCOP utilizando técnicas de inteligencia artificial para la detección de proveedores anómalos en los contratos.

Esto se logró obteniendo los datos de contratos de subasta inversa del sector salud realizados en el año 2020 por medio del SERCOP. A partir de su análisis se derivaron métricas de transparencia que reflejen el comportamiento de los proveedores al subastar. Se definió una arquitectura de modelo basado en GAAN. El entrenamiento y validación se hicieron iterativamente resultando en un grafo representativo de los proveedores y sus relaciones entre sí.

De 142 proveedores se obtuvieron 13 proveedores anómalos. Al analizar su comportamiento, se concluyó que existe una posibilidad en la cual estos proveedores cumplen un rol de relleno en sus contratos y coluden con otros proveedores para este objetivo.

Palabras Clave: SERCOP, detección de anomalías, GAAN, grafo atribuido, corrupción

Abstract

The public procurement system in Ecuador, SERCOP, has encountered numerous challenges, including corruption and inefficiency, resulting in a significant loss of public funds. Despite efforts to enhance transparency and efficiency, irregularities and inconsistencies persist in the procurement processes. Considering these circumstances, a model for characterizing suppliers of goods and services within SERCOP is proposed, utilizing artificial intelligence techniques for detecting anomalous suppliers in contracts.

This was achieved by obtaining data from reverse auction contracts in the healthcare sector conducted by SERCOP in 2020. Through analysis, transparency metrics were derived to reflect suppliers' behavior during auctions. An architecture based on the Generative Adversarial Autoencoder Network (GAAN) was defined. Training and validation were carried out iteratively, resulting in a representative graph depicting supplier relationships.

Among 142 suppliers, 13 anomalous suppliers were identified. Upon analyzing their behavior, it was concluded that there is a possibility these suppliers serve as fillers in contracts and collude with others to achieve this objective.

Keywords: *SERCOP, anomaly detection, GAAN, attributed graph, corruption*

CONTENIDO

Resumen	8
Abstract	9
Capítulo 1	15
1. Introducción	15
1.1 Descripción del problema.....	15
1.2 Justificación del problema	16
1.3 Objetivos.....	17
1.3.1 Objetivo general	17
1.3.2 Objetivos específicos.....	17
1.4 Marco teórico.....	17
1.4.1 Trabajos Relacionados	17
1.4.2 Bases Teóricas.....	19
Capítulo 2.....	24
2. Metodología.	24
2.1 Recolección de datos	25
2.2 Generación de Características de Transparencia	25
2.2.1 Métricas contratos perdidos	26
2.2.2 Métricas contratos ganados	31
2.3 Análisis y preparación de datos	37
2.4 Definición del modelo	37
2.4.1 DOMINANT	37
2.4.2 GAAN	38
2.5 Entrenamiento y Validación	40
2.5.1 Entrenamiento	41
2.5.2 Validación	42
2.6 Construcción del grafo que modela las relaciones entre los proveedores	43

2.7	Interpretación de los resultados	44
2.7.1	Identificación de Proveedores Anómalos.....	44
2.7.2	Consideraciones éticas y legales	45
Capítulo 3	46
3.	Resultados y análisis	46
3.1	Resultados del Modelo	46
3.1.1	Análisis de información estructural.....	46
3.1.2	Análisis de métricas para contratos perdidos	47
3.1.3	Análisis de métricas para contratos ganados.....	52
3.1.4	Análisis de caso excepción – proveedores de participación única.....	56
Capítulo 4	58
4.	Conclusiones y recomendaciones.....	58
4.1	Conclusiones.....	58
4.2	Recomendaciones	58
5.	Referencias	60

Abreviaturas

SERCOP Servicio Nacional de Contratación Pública

ESPOL Escuela Superior Politécnica del Litoral

SIE Subasta Inversa Electrónica

OCDS Open Contracting Data Standard

AE Auto-Encoder

GCN Graph Convolutional Network

GAN Generative Adversarial Network

Índice de figuras

Figura 1	20
Figura 2	21
Figura 3	24
Figura 4	37
Figura 5	39
Figura 6	44

Índice de tabla

Tabla 1.....	40
Tabla 2.....	41
Tabla 3.....	41
Tabla 4.....	42
Tabla 5.....	44
Tabla 6.....	46
Tabla 7.....	47
Tabla 8.....	47
Tabla 9.....	49
Tabla 10.....	50
Tabla 11.....	52
Tabla 12.....	54
Tabla 13.....	56

Capítulo 1

1. Introducción

1.1 Descripción del problema

La contratación y adquisiciones públicas son clave en la entrega de servicios públicos, la promoción de una gobernanza efectiva y el fomento de una economía sustentable. Ecuador posee un sistema de compras públicas llamado Sistema de Contratación Pública o SERCOP (Servicio Nacional de Contratación Pública). El sistema fue creado en el año 2007 con el objetivo de manejar los contratos públicos a nivel Nacional en un sistema informático que contribuya a la eficiencia y transparencia de los procesos de contratación pública. [1]

“Con el agotamiento del modelo de creciente gasto público, desde 2015, pero con más fuerza desde la pandemia, se ha vuelto más visible lo engorroso y poco eficiente que es el sistema de contratación pública”. [2] El SERCOP no cumple sus objetivos planteados debido a las irregularidades e inconsistencias que existen en los procesos de contratación y subastas inversas realizados por medio del sistema. La preocupación de los ecuatorianos por este tema es significativa por las grandes cantidades de dinero que se manejan en estos contratos.

En los primeros meses de pandemia en Ecuador [marzo, abril, mayo del 2020] se destaparon casos de corrupción que incluían sobreprecio en bolsas de cadáveres, mascarillas y pruebas de detección del covid-19, que alcanzaron 202 procesos irregulares identificados por el Servicio de Contratación Pública (SERCOP). El presidente de Ecuador, Guillermo Lasso, aseguró este lunes [13 de septiembre del 2021] en la suscripción de un acuerdo con la ONU para combatir la corrupción, que este "daño permanente y constante" le ha costado al país 70.000 millones de dólares en los últimos 14 años.

“Durante el 2021, las adquisiciones públicas sumaron a un total de USD 5.320 millones, llevadas a cabo por más de 3.600 entidades contratantes, que comprenden Gobiernos Autónomos Descentralizados (GADs), entidades de distintas funciones del Estado y empresas públicas. Con la cantidad de procedimientos realizados durante cada año, Jijón [directora del SERCOP] mencionó que es difícil determinar un porcentaje preciso de procesos con irregularidades” [3]

1.2 Justificación del problema

Una administración de gobierno efectiva depende en gran medida del manejo eficiente y transparente de las compras públicas. Este proceso es de vital importancia para garantizar el uso apropiado de los bienes públicos, fomentar la confianza de los ciudadanos en los organismos estatales y promover la integridad del gobierno. Sin embargo, una gestión cuidadosa y transparente de este sistema representa desafíos significativos, lo que genera la necesidad de una gestión transparentes y cuidadosa.

El estudio realizado por la Red Internacional de Compras Gubernamentales (RICG) y la Universidad de los Andes en 2021, mostró el valor de los datos abiertos de contratación para evaluar y mejorar los sistemas de compras públicas. Estos datos, al ser analizados con técnicas de aprendizaje automático y estadística, pueden revelar patrones importantes, permitir la detección de anomalías, etc.

Estas técnicas tienen mucho potencial cuando se utilizan para caracterizar a los proveedores en un sistema de compras públicas. Se puede mejorar la capacidad de detectar comportamientos inusuales entre los proveedores y proporcionar información útil a las autoridades reguladoras para que puedan dirigir mejor los esfuerzos de supervisión y regulación.

Es fundamental intensificar los mecanismos de control para detectar irregularidades en los procesos de adquisición pública. Con este objetivo, se deben desarrollar herramientas avanzadas, que mejoran la capacidad de reconocer comportamientos atípicos entre los diferentes proveedores de bienes y servicios. Además, al identificar patrones de comportamiento anómalos y localizar las

áreas en las que estos ocurren, se puede suministrar a las autoridades reguladoras información de gran valor. Este conocimiento permite orientar los esfuerzos de regulación y supervisión de manera más efectiva, lo que lleva a una gestión más eficiente y responsable.

En resumen, la propuesta de investigación para la caracterización de proveedores en un sistema de compras públicas tiene un gran potencial para mejorar la transparencia, la equidad y la eficiencia de los procesos de adquisición pública, reforzando así la confianza en la gestión gubernamental. Es, sin duda, una iniciativa de relevancia crítica para una administración de gobierno efectiva y responsable.

1.3 Objetivos

1.3.1 Objetivo general

Desarrollar un modelo de caracterización de los proveedores de bienes y servicios del sistema de compras públicas SERCOP utilizando técnicas de inteligencia artificial para la detección de proveedores anómalos en los contratos.

1.3.2 Objetivos específicos

1. Extraer características significativas de los proveedores de bienes y servicios en base a la información de los procesos de adquisiciones públicas que facilite su descripción detallada.
2. Desarrollar un modelo que identifique proveedores anómalos de bienes y servicios de un sistema de compras públicas, utilizando técnicas de aprendizaje automático.
3. Crear visualizaciones de los proveedores anómalos que participan en el proceso de contratación que facilite su análisis a organismos de control.

1.4 Marco teórico

1.4.1 Trabajos Relacionados

Se han realizado varios estudios para tratar de detectar anomalías o irregularidades en los procesos de adquisición de compras públicas. Por ejemplo:

M. E. K. Niessen et al. [4] Utilizaron los datos públicos de las adquisiciones de Paraguay, donde empleaban un modelo de aprendizaje automático para la detección de anomalías, basado en el algoritmo de Isolation Forest, lo notable de este enfoque es que permitía replicar la técnica desarrollada en otros países que implementan el mismo estándar de datos. Este estudio también permite asignar un puntaje a los contratos y procesos de adquisición, lo cual puede usarse para identificar anomalías, brindando un apoyo útil a las entidades gubernamentales de regulación y control de los procesos. Además, demostró que el modelo usado podía detectar con una precisión por encima del 90%. La alta tasa de éxito indica que el uso de metodologías de aprendizaje automático, como lo es el Isolation Forest en este caso, pueden ser una estrategia viable para la detección de irregularidades en las compras públicas.

Este enfoque no es óptimo para el caso del SERCOP debido a que hay diferencias significativas en la estructura y cantidad de datos en comparación al sistema de compras públicas de Paraguay. En el estudio se realiza un análisis en dos grupos de datos que corresponden a dos estados distintos involucrados en el proceso de contratación lo cual no es aplicable en los datos obtenidos para este caso de estudio ya que es enfocado a los contratos realizados por medio de subasta inversa solamente. Además, se utilizan técnicas de aprendizaje no supervisado de Machine Learning (ML) las cuales no pueden realizar un análisis en representaciones de redes a diferencia de técnicas de Deep Learning (DL) como las que se usaron en este caso de estudio.

Wang, X. et al. [5] Propone una red neuronal de grafos con una sola clase (OCGNN, por sus siglas en inglés), para la detección de anomalías en grafos. Este modelo está diseñado para combinar toda la capacidad de representación de las redes neuronales de grafos con el objetivo de una sola clase. En comparación con otros modelos, este logra mejoras importantes en pruebas con gran cantidad de datos.

Sin embargo, en este caso de estudio no es óptima esta estrategia puesto que se necesita conocer y agrupar con exactitud los nodos que presentan anomalías para luego poder aplicar esta

información en una nueva visualización del grafo. Es decir, la información sobre las relaciones de los nodos anómalos debe mantenerse y, con la estrategia de obtener los nodos aberrantes que caen fuera de la esfera no anómala, se pierde.

X. Ma et al. [6] Propone una revisión esquemática y completa de las diferentes técnicas de aprendizaje profundo para la detección de anomalías en grafos. A manera detallada, su enfoque radica en los desafíos de esta área de investigación y discute sus fundamentos claves, las especificaciones técnicas, las ventajas y desventajas relativas de varias técnicas. Los resultados, destacan un sin número de direcciones de investigación futura que acaparan problemas no resueltos y emergentes introducidos por los grafos, tales como: detección de anomalías, aprendizaje profundo, etc.

El estudio destaca tres tipos de anomalías que pueden ser identificadas en representaciones de grafos o redes: globales, estructurales y comunidades. Lo ideal para este caso de estudio es llegar a identificar anomalías de tipo comunidad, ya que relacionan la información de los nodos anómalos con las relaciones (arcos) anómalas que existen en el grafo para destacar grupos de nodos relacionados (comunidades) dentro del mismo. También añade que el uso de técnicas de DL para detectar este tipo de anomalías ha sido explorado en pocos trabajos de investigación hasta la fecha, pero tienen gran potencial para obtener resultados disruptivos.

1.4.2 Bases Teóricas

1.4.2.1 Servicio Nacional de Compras Publicas

Es el ente que establece las condiciones y pautas para las compras del gobierno a nivel nacional. Sus procesos de contratación se basan en principios de transparencia, calidad, eficiencia, efectividad y ahorro de recursos para el gobierno. Además de que se basan en el artículo 288 de la constitución del Ecuador, donde se menciona que se debe dar

preferencia a los productos y servicios locales, en especial a los que proceden de la economía social y solidaria y de las micro, pequeñas y medianas empresas. [7]

1.4.2.2 OCDS

El Estándar de Datos de Contrataciones Abiertas (OCDS) es un estándar público y de libre uso para la información de contrataciones públicas, adoptado por más de 30 gobiernos a nivel mundial. El estándar establece las directrices para la publicación de datos y documentos vinculados a los procesos de contratación de bienes y servicios [8].

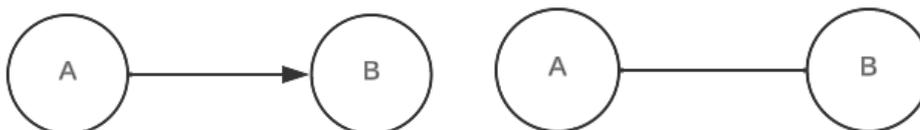
El objetivo principal del OCDS es garantizar que las contrataciones abiertas se divulguen utilizando datos abiertos uniformes. Además, el OCDS proporciona instrucciones sobre cómo publicar información y documentos relacionados con la adquisición de bienes, obras y servicios. Se encarga de que los datos de contratación sean accesibles para que cualquier persona pueda utilizarlos, alterarlos y compartirlos para cualquier finalidad. [9]

1.4.2.3 Grafos Atribuidos

Un grafo es un tipo de dato abstracto (TDA) conformado por vértices o nodos y conexiones entre ellos llamadas aristas o arcos. Las aristas pueden ser bidireccionales o unidireccionales, determinando la dirección de la relación entre los vértices. A un grafo bidireccional, es decir, un grafo en el cual todas sus aristas son bidireccionales, también se le conoce como no dirigido.

Figura 1

Dos ejemplos de grafos con aristas unidireccionales (izquierda) y bidireccionales (derecha)



Nota: La imagen es de autoría propia

Un grafo es definido como $G = (V, E)$ donde V es un conjunto finito $V = \{1, 2, 3, 4, \dots, N\}$ de vértices y E es un conjunto de aristas. Si este nodo tiene atributos este grafo pasa a ser un grafo atribuido definido como $G = (V, E, l)$ donde l es una función de etiquetas $l: V \rightarrow LN$ [1]

Se define al Grado de un vértice V , $\text{grado}(V)$, en grafos no dirigidos como el número de aristas incidentes en V [1]. Sea $G = (V, E)$ un grafo con m aristas:

$$\sum_{V \in G} \text{grado}(V) = 2m$$

Existen varias representaciones de un grafo, una de estas es la llamada matriz de adyacencia o booleana.

Sea $G = (V, E)$ un grafo no dirigido con $|V| = n$. Se define la matriz de adyacencia o booleana asociada a G como $B_{n \times n}$ con:

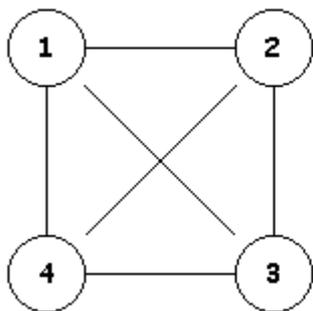
$$b_{i,j} = \begin{cases} 1, & (i,j) \in E \\ 0, & (i,j) \notin E \end{cases}$$

La matriz B es simétrica con 1 en las posiciones (i, j) y (j, i) si existe la arista (i, j) .

Un ejemplo se muestra en la Figura 2

Figura 2

Ejemplo de grafo no dirigido y su matriz de adyacencia correspondiente



	1	2	3	4
1	1	1	1	1
2	0	0	1	1
3	0	0	0	1
4	0	0	0	0

Matriz de Adyacencia del Grafo.

Nota: Imagen obtenida de [2]

1.4.2.4 Distribución Gaussiana

La distribución Gaussiana es una distribución de probabilidad continua que tiene simetría alrededor de su valor medio, la mayoría de las observaciones se concentran cerca del valor central y las probabilidades de valores que se alejan significativamente del valor medio decrecen de manera igualitaria en ambas direcciones. A pesar de que la distribución Gaussiana es simétrica, no todas las distribuciones simétricas son gaussianas. [3]

La función de densidad de probabilidad de la distribución normal es:

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

x es la variable

μ es el promedio

σ es la desviación estandar

1.4.2.5 Auto-Encoder o AE

El auto-encoder es una red neuronal no supervisada que aprende a comprimir y codificar los datos de forma eficaz y, a continuación, aprende a reconstruir los datos a partir de la representación codificada reducida a una representación lo más parecida posible a la entrada original. El auto-encoder por diseño, reduce las dimensiones de los datos aprendiendo a ignorar el ruido que contienen. [4]

1.4.2.6 Embeddings o incrustaciones

Un embedding o incrustación es una representación matricial de un objeto, como una palabra o imagen, en un espacio de múltiples dimensiones. Esto nos permite captar las relaciones entre los distintos objetos de un conjunto de datos y comprender cómo se relacionan entre sí. La dimensionalidad de un embedding define la cantidad de información que puede ser obtenida de un objeto. A mayor dimensionalidad, más información podrá ser recopilada. Sin embargo, esto incrementa la carga computacional. [5]

1.4.2.7 GCN

Una Graph Convolutional Network (GCN) es un enfoque de aprendizaje semisupervisado para datos estructurados en forma de grafos. Se fundamenta en una versión eficaz de las redes neuronales convolucionales que trabajan directamente con grafos. La selección de la arquitectura convolucional se basa en una aproximación local del primer orden de las convoluciones espectrales en grafos. El modelo tiene una escala lineal en relación con la cantidad de aristas en el grafo y aprende representaciones de la capa oculta que encapsulan tanto la estructura del grafo como las características de los nodos. [6]

1.4.2.8 GAN

Una red generativa adversarial (GAN) es un modelo de aprendizaje automático (ML) en el que dos redes neuronales compiten entre sí utilizando métodos de aprendizaje profundo para ser más precisas en sus predicciones. Las GAN suelen ejecutarse sin supervisión y utilizando un marco cooperativo de suma cero para aprender. Usan un generador, que aprende a producir la salida objetivo, con un discriminante, que aprende a distinguir datos verdaderos de la salida del generador. El generador intenta engañar al discriminador y el discriminante intenta evitar que se engañe. [7]

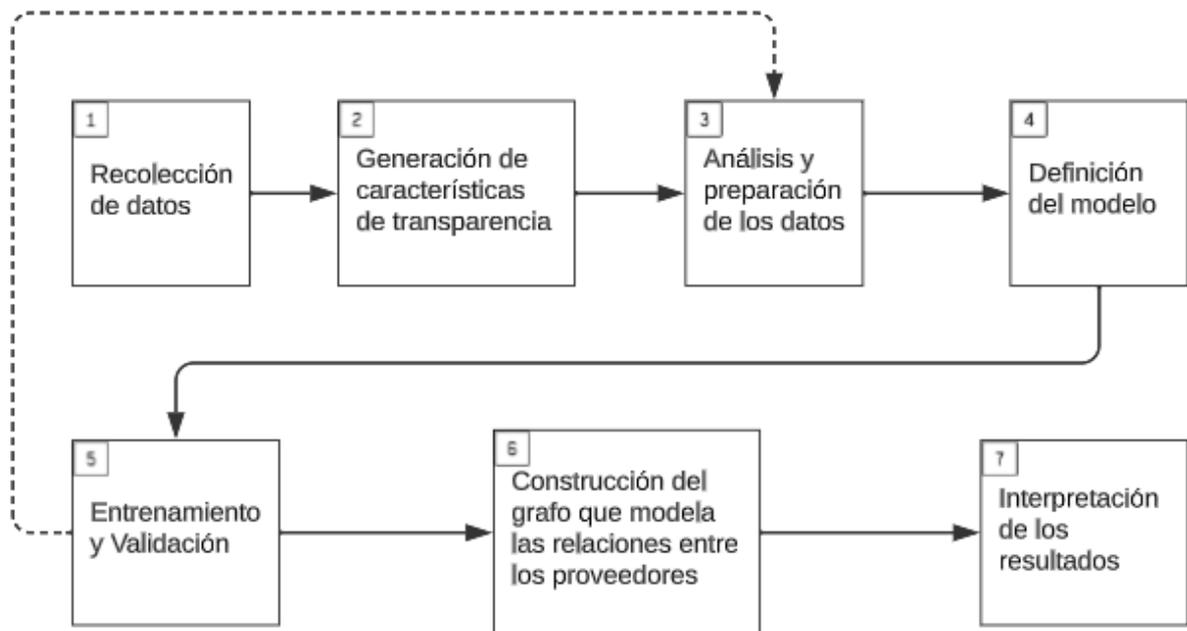
Capítulo 2

2. Metodología.

La metodología se dividió en 7 etapas de manera secuencial, es decir, lo que se realizó en la etapa 3 dependerá de lo que se desarrolló en la etapa 2, como se observa en la Figura 3. Las etapas son: Recolección de datos, generación de características/métricas de transparencia, análisis y preparación de los datos, definición del modelo, entrenamiento y validación, construcción del grafo que modela las relaciones entre los proveedores e interpretación de los resultados.

Figura 3

Metodología



Nota: La imagen es de autoría propia

2.1 Recolección de datos

Durante esta etapa se utilizó una base de datos previamente creada por el grupo anterior, la cual contenía información de todos los contratos del SERCOP desde el año 2018. Dado que el enfoque principal de este estudio se centra en el ámbito de la salud, se recopilaron exclusivamente los contratos de tipo Subasta Inversa Electrónica (SIE) relacionados con solo este sector en el año 2020.

Si bien la información obtenida, resultó vital para el inicio de este estudio, se identificó la necesidad de recopilar nuevamente datos adicionales para la generación de métricas relevantes. Específicamente, se necesitaba información adicional, como el monto inicial del contrato, las diferentes ofertas realizadas por los diversos proveedores. Esta información adicional es crucial porque permite obtener una visión más completa y detallada de la situación, lo cual es esencial para realizar un análisis más exhaustivo y riguroso en la transparencia de los contratos en el sector salud. Al recopilar estos nuevos datos, se puede evaluar de manera más precisa las relaciones entre los proveedores, identificar patrones y tendencias.

Por lo tanto, se procedió a realizar una nueva recopilación de información de los contratos de este sector con la obtención de los datos faltantes para lograr una caracterización completa y con más detalle de los proveedores.

2.2 Generación de Características de Transparencia

Una vez finalizada la recolección de los datos, la siguiente etapa fue la generación de las métricas, cuyo principal propósito era caracterizar de forma cuantitativa a los proveedores. La creación de estas métricas estuvo precedida de una meticulosa definición de conceptos esenciales que sentaron las bases para un mejor entendimiento profundo y claro de las mismas.

Primeramente, se estableció el concepto de *“Participación en un contrato”*. Se conceptualizó como aquel escenario en el cual un proveedor había presentado, al menos, una oferta en la subasta. Un contrato en el que el proveedor no realizaba ninguna oferta quedaba excluido de la consideración de participación.

Posteriormente, se delinearón dos categorías de contratos basados en la naturaleza de la participación del proveedor en la subasta: “*Contratos únicos o de participación única(U)*” y “*Contratos colectivos o de participación colectiva (C)*”. Los primeros se identificaron como aquellos en los que el proveedor era el único en presentar una oferta, obteniendo así el contrato sin tener competencia alguna. En cambio, los contratos colectivos se definieron como aquellos en los que, a pesar de la presencia de múltiples ofertas de distintos proveedores, el proveedor en cuestión lograba superar a sus competidores y adjudicarse el contrato.

Además, se contempló la categoría de “*Contratos perdidos (P)*”, refiriéndose a aquellos contratos en los que el proveedor, pese a haber realizado una oferta en la subasta, no resultaba ser el ganador. Finalmente, también se definieron los conceptos de “*Monto Referencial*” y “*Monto Ganador*”, el primero de ellos hace referencia al monto inicial del contrato el cual es estipulado y definido por la entidad que ofrece el contrato. Por otro lado, monto ganador hace referencia al monto por el cual el contrato fue adjudicado a un proveedor.

Con estas definiciones fundamentales establecidas, se procedió a calcular una serie de métricas pertinentes, que ofrecerían una visión más estructurada y analítica del comportamiento y desempeño de los proveedores en su historial de participación en contratos públicos. A continuación, se presenta el listado de métricas segmentado basado en el tipo de contrato:

2.2.1 Métricas contratos perdidos

Promedio de Ofertas en Contratos Perdidos (POCP): Esta métrica toma en cuenta el promedio total de los montos ofertados por el proveedor en cada contrato. Estos montos se normalizan en relación con el monto de referencia del contrato correspondiente. Además, se considera la cantidad total de ofertas realizadas durante las subastas en los contratos que ha pedido.

Un valor alto indica que, en promedio las ofertas guardan una estrecha correlación con el monto referencial del contrato; mientras que un valor bajo indica que, en promedio, las ofertas guardan una amplia correlación con el monto referencial del contrato.

$$\mu_{ofertas\ perdidias} = \frac{1}{\sum_{j=1}^m n_j} \sum_{j=1}^m \sum_{i=1}^{n_j} \frac{O_{ij}}{R_j}$$

m: Representa el número total de contratos en los que el proveedor ha participado y ha perdido.

n_j: Representa el número de ofertas realizadas por el proveedor en el contrato

O_{ij}: Representa la oferta *i* realizada por el proveedor en el contrato *j*

R_j: Representa el monto referencial del contrato *j*.

Porcentaje de Contratos Perdidos por proveedor (PCP): Esta métrica calcula el porcentaje de contratos perdidos en relación con el total de contratos en los que el proveedor ha participado. Ofrece una medida relativa de los contratos que el proveedor no ha podido obtener.

Un valor alto indica que la mayoría de los contratos en los que el proveedor ha participado los ha perdido.

$$\%Cp = \frac{N(P)}{N(U) + N(C) + N(P)}$$

N(U): Representa la cantidad de contratos de participación única

N(C): Representa la cantidad de contratos de participacion colectiva

N(P): Representa la cantidad de contratos perdidos

Desviación Estándar Promedio de los Contratos Perdidos (DEPCP): Esta métrica es el promedio de las desviaciones estándar (normalizadas contra el monto referencial del contrato) de todas las ofertas realizadas en todos los contratos perdidos por el proveedor. Esta medida nos muestra qué tanto se diferencian las ofertas que realiza un proveedor en una subasta.

Un valor alto indica una estrecha correlación entre todas las ofertas de los contratos que ha perdido un proveedor. Un valor bajo indica una amplia correlación entre todas las ofertas de los contratos que ha perdido un proveedor

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\mu_{desviacion\ estandar\ perdidos} = \frac{1}{m} \sum_{j=1}^m \sigma_j$$

$$\mu_{desviacion\ estandar\ perdidos} = \frac{1}{m} \sum_{j=1}^m \sqrt{\frac{1}{n_j} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}$$

m: Representa el numero de contratos en los que el proveedor participó y perdió

n_j: Es el numero de ofertas en el contrato *j*

x_{ji}: Es la oferta *i* en el contrato *j*

x̄_j: Es el promedio de las ofertas en el contrato *j*

Promedio de la Diferencia de la Última Oferta con respecto al Monto Ganador en Contratos Perdidos (PDUOMGCP): Esta métrica compara la última oferta realizada por un proveedor con el monto ganador del contrato. Permite evaluar si el proveedor estuvo cerca o lejos de obtener el contrato en función de su oferta final y proporciona información sobre su nivel de competitividad en relación con otros proveedores.

Un valor alto indica que, en promedio, las ofertas finales del proveedor en sus contratos perdidos están alejadas del monto ganador de sus contratos y una baja competitividad por parte del proveedor.

$$D_j = \frac{|G_j - O_j|}{G_j}$$

$$\mu_{diferencia\ ganador} = \frac{1}{m} \sum_{j=1}^m D_j$$

m: Representa el numero de contratos en los que el proveedor participo y ha perdido

O_j: Es la última oferta realizada por el proveedor en el contrato *j*

G_j: Es el monto ganador del contrato *j*

Promedio de la diferencia de la última oferta con respecto al monto planeado en contratos perdidos (PDUOMPCP): Esta métrica evalúa la diferencia entre la última oferta realizada por un proveedor y el monto planeado inicialmente para el contrato. Permite analizar si el proveedor se ajusta o se desvía del monto estimado y proporciona información sobre su capacidad para ofrecer precios competitivos.

Un valor alto indica que, en promedio, las ofertas finales del proveedor en sus contratos perdidos están alejadas del monto referencial de sus contratos y una alta competitividad por parte del proveedor.

$$D_j = \frac{|R_j - O_j|}{R_j}$$

$$\mu_{diferencia\ planeado} = \frac{1}{m} \sum_{j=1}^m D_j$$

m: Representa el número de contratos en los que el proveedor participó y perdió

O_j: Es la última oferta realizada por el proveedor en el contrato *j*

R_j: Es el monto referencial del contrato *j*

Promedio de Frecuencia de Contratos Perdidos (PFCP): Esta métrica indica el promedio de la frecuencia de las ofertas de un proveedor en los contratos que ha perdido.

Un valor alto indica que el proveedor realizó varias ofertas con montos alejados entre sí en comparación a los otros proveedores ofertantes en sus contratos. Un valor bajo indica que el proveedor realizó pocas ofertas con montos cercanos entre sí en comparación a los otros proveedores ofertantes en sus contratos.

$$Frecuencia = \frac{\sigma_{ofertas\ proveedor}}{\sum \sigma_{ofertas\ competidores}}$$

$$\mu_{frecuencia\ cp} = \frac{\sum frecuencia_{cp}}{CP}$$

Rango de Índice de Clasificación en Contratos Perdidos (RICCP): Resta del índice de clasificación máximo y mínimo en los contratos perdidos. Nos indica la variedad de cuántos proveedores clasifican a la subasta en los contratos perdidos de un proveedor.

Un valor alto indica que el proveedor ha participado tanto en contratos donde la mayoría de los proveedores ha clasificado como en contratos donde pocos proveedores han clasificado. Un valor bajo indica que el proveedor ha participado solamente en contratos donde la mayoría de los proveedores ha clasificado o solamente en contratos donde pocos proveedores han clasificado.

$$g(x, y) = \begin{cases} 1, & y = x \\ \left| \log_{10} \left(\frac{x}{y} \right) \right|, & y > x \end{cases}$$

$$\text{Índice_rango} = \max(g_1, g_2, g_3, \dots, g_n) - \min(g_1, g_2, g_3, \dots, g_n)$$

$g(x, y)$: Representa la función de Índice de clasificación

g_i : Representa el índice de clasificación para el contrato i

x : Representa la cantidad de proveedores que participaron en la subasta

y : Representa la cantidad de proveedores que se inscribieron en el contrato

Promedio de Índice de Clasificación en Contratos Perdidos (PICCP): Promedio del índice de clasificación de un proveedor en los contratos perdidos. Nos indica el valor promedio de cuántos proveedores clasifican a la subasta en los contratos que un proveedor ha perdido.

Un valor alto indica que el proveedor ha participado en su mayoría, en contratos donde pocos proveedores han clasificado. Un valor bajo indica que el proveedor ha participado en su mayoría, en contratos donde la mayoría de los proveedores ha clasificado.

$$g(x, y) = \begin{cases} 1, & y = x \\ \left| \log_{10} \left(\frac{x}{y} \right) \right|, & y > x \end{cases}$$

$$\text{mean_indice} = \frac{\sum_i^n g_i}{n}$$

$g(x, y)$: Representa la función de Índice de clasificación

g_i : Representa el índice de clasificación para el contrato i

x : Representa la cantidad de proveedores que participaron en la subasta

y : Representa la cantidad de proveedores que se inscribieron en el contrato

2.2.2 Métricas contratos ganados

Promedio de Ofertas en Contratos Colectivos (POCC): Esta métrica toma en cuenta el promedio total de los montos ofertados por el proveedor en cada contrato. Estos montos se normalizan en relación con el monto de referencia del contrato correspondiente. Además, se considera la cantidad total de ofertas realizadas durante las subastas en los contratos de participación colectiva.

Un valor alto indica que, en promedio las ofertas guardan una estrecha correlación con el monto referencial del contrato; mientras que un valor bajo indica que, en promedio, las ofertas guardan una amplia correlación con el monto referencial del contrato.

$$\mu_{\text{ofertas colectivas}} = \frac{1}{\sum_{j=1}^m n_j} \sum_{j=1}^m \sum_{i=1}^{n_j} \frac{O_{ij}}{R_j}$$

m : Representa el número total de contratos colectivos de un proveedor.

n_j : Representa el número de ofertas realizadas por el proveedor en el contrato

O_{ij} : Representa la oferta i realizada por el proveedor en el contrato j

R_j : Representa el monto referencial del contrato j .

Porcentaje de Contratos Colectivos por Proveedor (PCC): Esta métrica se utiliza para analizar el éxito de un proveedor en la obtención de contratos. Se calcula evaluando la proporción

de contratos que el proveedor ha ganado al participar de manera colectiva en comparación con el total de contratos en los que el proveedor ha participado.

Un valor alto de este porcentaje indica que la mayoría de los contratos obtenidos por el proveedor fueron logrados al participar en contratos colectivos.

$$\%Cc = \frac{N(C)}{N(U) + N(C) + N(P)}$$

N(U): Representa la cantidad de contratos de participacion unica

N(C): Representa la cantidad de contratos de participacion colectiva

N(P): Representa la cantidad de contratos perdidos

Promedio Monto Ganado en Contratos Colectivos (PMGCC): Esta métrica refleja la media de los montos de las ofertas ganadores en las subastas en las que el proveedor participo de forma colectiva. Un valor alto en el monto promedio indica que el proveedor ha obtenido contratos con valores referenciales elevados. Esto permitiría identificar si un proveedor suele ganar contratos con montos altos o viceversa.

$$\mu_{\text{monto ganado colectivos}} = \frac{\sum_{i=1}^m G_i}{m}$$

m: Representa el número total de contratos colectivos en los que el proveedor ha ganado la subasta

G_i: Es la oferta ganadora realizada por el proveedor en el contrato j

Desviación Estándar Promedio de los Contratos Colectivos (DEPCC): Esta métrica calcula el promedio de las desviaciones estándar normalizadas de las ofertas presentadas en cada contrato en el que el proveedor participó de forma colectiva, contra la cantidad de contratos. Un valor alto en esta métrica indica que el proveedor suele ofertar valores bastante alejados entre sí.

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\mu_{\text{desviacion estandar colectivos}} = \frac{1}{m} \sum_{j=1}^m \sigma_j$$

$$\mu_{desviacion\ estandar\ colectivos} = \frac{1}{m} \sum_{j=1}^m \sqrt{\frac{1}{n_j} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}$$

m: Representa el número de contratos en los que el proveedor participó de forma colectiva

n_j: Es el número de ofertas en el contrato *j*

x_{ji}: Es la oferta *i* en el contrato *j*

x̄_j: Es el promedio de las ofertas en el contrato *j*

Promedio de la Diferencia de la Última Oferta con respecto al monto planeado en Contratos Colectivos (PDUOCC): Esta métrica evalúa la diferencia entre la última oferta realizada por un proveedor y el monto planeado inicialmente para el contrato. Permite analizar si el proveedor se ajusta o se desvía del monto estimado y proporciona información sobre su capacidad para ofrecer precios competitivos.

Un valor alto indica que, en promedio, las ofertas finales del proveedor en sus contratos perdidos están alejadas del monto referencial de sus contratos y una alta competitividad por parte del proveedor.

$$D_j = \frac{|R_j - O_j|}{R_j}$$

$$\mu_{diferencia\ planeado\ únicos} = \frac{1}{m} \sum_{j=1}^m D_j$$

m: Representa el número de contratos en los que el proveedor participo de forma colectiva

O_j: Es la última oferta realizada por el proveedor en el contrato *j*

R_j: Es el monto referencial del contrato *j*

Promedio de frecuencia de contratos perdidos (PFCP): Esta métrica indica el promedio de la frecuencia de las ofertas de un proveedor en los contratos que ha perdido. Un valor alto indica que el proveedor realizó varias ofertas con montos alejados entre sí en comparación a los otros proveedores ofertantes en sus contratos. Un valor bajo indica que el proveedor realizó pocas ofertas con montos cercanos entre sí en comparación a los otros proveedores ofertantes en sus contratos.

$$Frecuencia = \frac{\sigma_{ofertas\ proveedor}}{\sum \sigma_{ofertas\ competidores}}$$

$$\mu_{frecuencia\ cp} = \frac{\sum frecuencia_{cp}}{CP}$$

Rango de Índice de Clasificación en Contratos Colectivos (RICCC): Resta del índice de clasificación máximo y mínimo en los contratos colectivos. Nos indica la variedad de cuántos proveedores clasifican a la subasta en los contratos colectivos de un proveedor.

Un valor alto indica que el proveedor ha participado tanto en contratos donde la mayoría de los proveedores ha clasificado como en contratos donde pocos proveedores han clasificado. Un valor bajo indica que el proveedor ha participado solamente en contratos donde la mayoría de los proveedores ha clasificado o solamente en contratos donde pocos proveedores han clasificado.

$$g(x, y) = \begin{cases} 1, & y = x \\ \left| \log_{10} \left(\frac{x}{y} \right) \right|, & y > x \end{cases}$$

$$\text{Índice_rango} = \max(g_1, g_2, g_3, \dots, g_n) - \min(g_1, g_2, g_3, \dots, g_n)$$

$g(x, y)$: Representa la función de Índice de clasificación

g_n : Representa el índice de clasificación para el contrato n

x : Representa la cantidad de proveedores que participaron en la subasta

y : Representa la cantidad de proveedores que se inscribieron en el contrato

Promedio de Índice de Clasificación en Contratos Colectivos (PICCC): Promedio del índice de clasificación de un proveedor en los contratos colectivos. Nos indica el valor promedio de cuántos proveedores clasifican a la subasta en los contratos que un proveedor ha participado de manera colectiva.

Un valor alto indica que el proveedor ha participado en su mayoría, en contratos donde pocos proveedores han clasificado. Un valor bajo indica que el proveedor ha participado en su mayoría, en contratos donde la mayoría de los proveedores ha clasificado.

$$g(x, y) = \begin{cases} 1, & y = x \\ \left| \log_{10} \left(\frac{x}{y} \right) \right|, & y > x \end{cases}$$

$$mean_indice = \frac{\sum_i^n g_i}{n}$$

$g(x, y)$: Representa la función de Índice de clasificación

g_i : Representa el índice de clasificación para el contrato i

x : Representa la cantidad de proveedores que participaron en la subasta

y : Representa la cantidad de proveedores que se inscribieron en el contrato

2.2.2.1 Métricas contratos únicos

Porcentaje de Contratos Únicos (PCU): Esta métrica se utiliza para analizar el éxito de un proveedor en la obtención de contratos. Se calcula evaluando la proporción de contratos que el proveedor ha ganado al participar de manera única en comparación con el total de contratos en los que el proveedor ha participado.

Un valor alto de este porcentaje indica que la mayoría de los contratos obtenidos por el proveedor fueron logrados al participar de manera única.

$$\%Cu = \frac{N(U)}{N(U) + N(C) + N(P)}$$

$N(U)$: Representa la cantidad de contratos de participacion unica

$N(C)$: Representa la cantidad de contratos de participacion colectiva

$N(P)$: Representa la cantidad de contratos perdidos

Promedio de Índice de Clasificación en Contratos Únicos (PICCU): Promedio del índice de clasificación de un proveedor en los contratos únicos. Nos indica el valor promedio de cuantos proveedores clasifican a la subasta en los contratos que un proveedor ha participado de manera única.

Un valor alto indica que el proveedor ha participado en su mayoría, en contratos donde pocos proveedores han clasificado. Un valor bajo indica que el proveedor ha participado en su mayoría, en contratos donde la mayoría de los proveedores ha clasificado.

$$g(x, y) = \begin{cases} 1, & y = x \\ \left| \log_{10} \left(\frac{x}{y} \right) \right|, & y > x \end{cases}$$

$$\text{mean_indice} = \frac{\sum_i^n g_i}{n}$$

$g(x, y)$: Representa la función de Índice de clasificación

g_i : Representa el índice de clasificación para el contrato i

x : Representa la cantidad de proveedores que participaron en la subasta

y : Representa la cantidad de proveedores que se inscribieron en el contrato

Rango de Índice de Clasificación en Contratos Únicos (RICCU): Resta del índice de clasificación máximo y mínimo en los contratos únicos. Nos indica la variedad de cuántos proveedores clasifican a la subasta en los contratos únicos de un proveedor.

Un valor alto indica que el proveedor ha participado tanto en contratos donde la mayoría de los proveedores ha clasificado como en contratos donde pocos proveedores han clasificado. Un valor bajo indica que el proveedor ha participado solamente en contratos donde la mayoría de los proveedores ha clasificado o solamente en contratos donde pocos proveedores han clasificado.

$$g(x, y) = \begin{cases} 1, & y = x \\ \left| \log_{10} \left(\frac{x}{y} \right) \right|, & y > x \end{cases}$$

$$\text{Índice_rango} = \max(g_1, g_2, g_3, \dots, g_n) - \min(g_1, g_2, g_3, \dots, g_n)$$

$g(x, y)$: Representa la función de Índice de clasificación

g_n : Representa el índice de clasificación para el contrato n

x : Representa la cantidad de proveedores que participaron en la subasta

y : Representa la cantidad de proveedores que se inscribieron en el contrato

2.3 Análisis y preparación de datos

Se procedió a realizar un análisis detallado de los datos recopilados de los contratos del sector salud. Se calcularon y obtuvieron diferentes medidas descriptivas, tales como: promedios, medianas, desviaciones estándar y rangos para obtener estadísticas resumidas acerca de las métricas de caracterización. Asimismo, en esta etapa se realizó una selección de los datos, puesto que existían datos que no aportaban valores de manera significativa. Es importante señalar que este análisis no fue estático; basados en los hallazgos y retroalimentación que se obtuvieron en la etapa 2.5.2, se realizaron ajustes y mejoras en esta fase para optimizar el rendimiento del modelo.

2.4 Definición del modelo

Se realizó una investigación extensiva sobre las diferentes arquitecturas utilizadas para problemas de detección de anomalías en grafos atribuidos y se encontraron dos arquitecturas prometedoras para el contexto del problema a resolver: DOMINANT y GAAN.

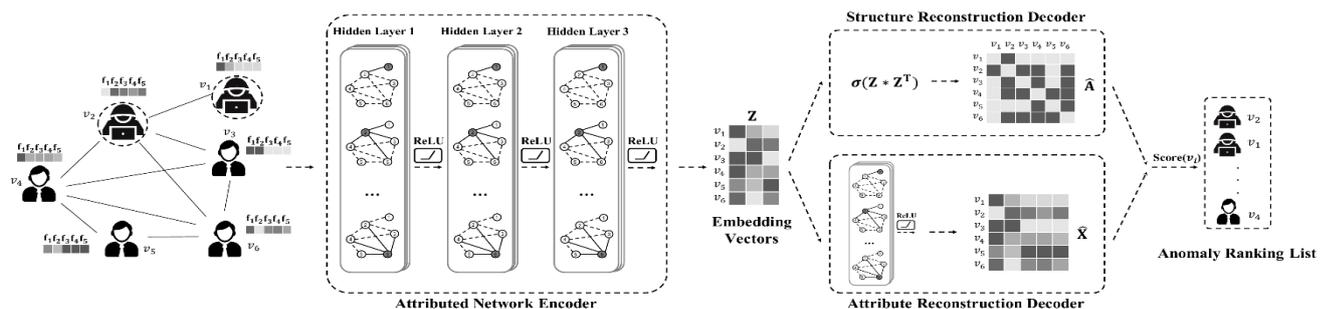
A continuación, se presenta en detalle la arquitectura de ambos modelos:

2.4.1 DOMINANT

DOMINANT es un modelo que busca realizar la tarea de detección de anomalías en Grafos Atribuidos por medio de la resolución de un problema de clasificación por ranking. Es decir, buscan clasificar todos los nodos según el grado de anormalidad, de manera que los nodos que difieren de manera singular de la mayoría de los nodos de referencia deberían ser clasificados en posiciones altas. La arquitectura del modelo se observa en la Figura 4.

Figura 4

Arquitectura del modelo DOMINANT



Nota: Obtenido de [8]

Como se puede observar, el bloque fundamental de construcción de DOMINANT es el Auto-Encoder o AE profundo, y consta de tres componentes esenciales:

1. Codificador de red atribuida, que modela la estructura de la red y los atributos nodales de manera conjunta para el aprendizaje de representaciones de Embeddings o incrustaciones de nodos con GCN
2. Decodificador de reconstrucción de estructura, cuyo objetivo es reconstruir la topología original de la red con las incrustaciones de nodos aprendidas.
3. Decodificador de reconstrucción de atributos, que intenta reconstruir los atributos nodales observados con las incrustaciones de nodos obtenidas.

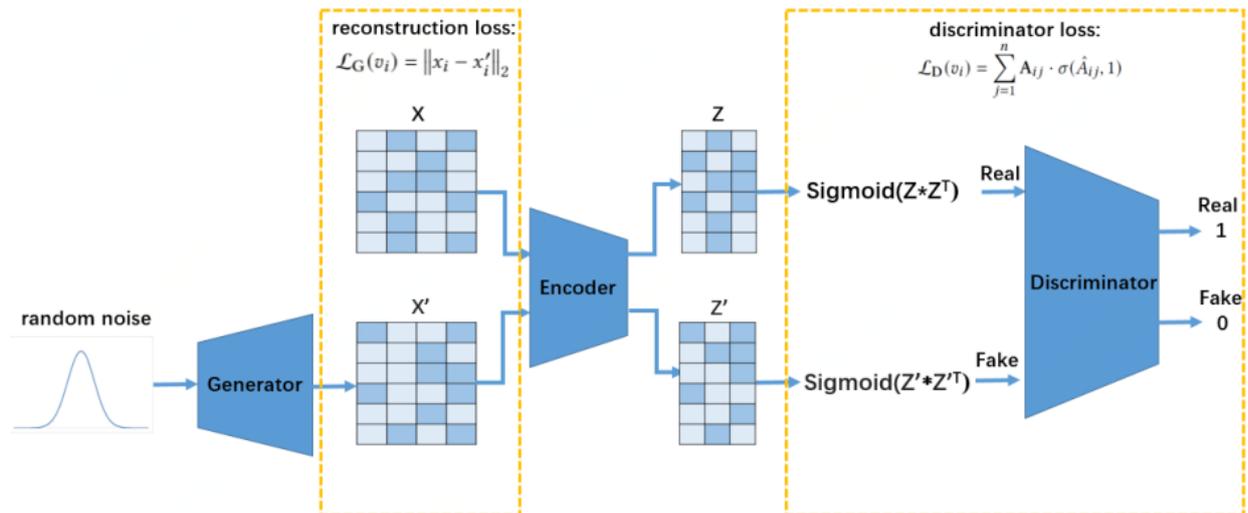
Posteriormente, los errores de reconstrucción de los nodos se utilizan para calcular un puntaje para cada nodo. Mientras más alto el puntaje, más anómalo es el nodo. En base a este puntaje se elabora un ranking donde los nodos en las posiciones más altas serían los de mayor grado de anomalía.

2.4.2 GAAN

Generative Adversarial Attributed Network (GAAN), es un modelo que busca realizar la tarea de detección de anomalías en grafos atribuidos basándose en el enfoque de las redes GAN: entrenar un generador G, el cual produce información falsa parecida a la real; y, a su vez, entrenar un discriminador D, el cual distingue la información falsa de la real. La arquitectura del modelo se presenta en la Figura 5.

Figura 5

Arquitectura del modelo GAAN



Nota: Obtenido de [9]

Como se observa, el modelo GAAN está compuesto por tres pasos o partes esenciales:

1. Los nodos de grafo falsos son generados por un módulo generador con ruido gaussiano como entrada. Se emplea un módulo codificador para mapear tanto los nodos de grafo reales como los falsos en un espacio latente.
2. Para codificar la información de la estructura del grafo en la representación latente de los nodos, calculamos la matriz de covarianza de muestra para los nodos reales y los nodos falsos, respectivamente.
3. Se entrena un discriminador para reconocer si dos nodos conectados provienen del grafo real o del grafo falso.

Con la salida del módulo codificador aprendido, se utiliza una medida de evaluación de anomalías que considera el error de reconstrucción de muestra y la confianza en la identificación de muestras reales para hacer predicciones.

2.5 Entrenamiento y Validación

Antes de proceder al entrenamiento y validación, fue necesario seleccionar el modelo adecuado para nuestro estudio. Elegimos el modelo GAAN sobre el modelo DOMINANT debido a la naturaleza de nuestro conjunto de datos.

Dado que nuestro conjunto de datos tiene un limitado número de nodos, fue complicado obtener datos anómalos sin sesgo en la estructura, además que la forma en que se encontraban las métricas de caracterización (valores de tipo decimal, pequeños) para DOMINANT fue complicado generalizar esto al momento de propagar el paso de mensajes por los nodos, lo que hizo que los nodos detectados como anómalos no compartieran valores similares en sus métricas.

Por otro lado, GAAN con su generador de ruido, nos permitió crear nodos muy parecidos a los nodos reales lo que facilitó al modelo aprende a discriminar de mejor forma si un nodo es o no anómalo. Esta característica fue crucial y de gran importancia para nuestro análisis, ya que buscamos detectar nodos anómalos a partir de sus métricas.

Para la implementación se optó por usar la librería PyGOD [10], una librería de Python para la detección de valores atípicos en grafos (detección de anomalías); incluye más de 10 algoritmos de detección y está desarrollada sobre PyTorch Geometric (PyG) [11] y PyTorch [12].

Entre los algoritmos disponibles poseen DOMINANT y GAAN. En la Tabla 1 se muestra un desglose de los hiperparámetros configurables en la implementación de PyGOD para el modelo GAAN.

Tabla 1

Hiperparámetros configurables para implementación de GAAN en PyGOD

HIPERPARÁMETRO	DESCRIPCIÓN
noise_dim	Dimensión del ruido gaussiano
hid_dim	Dimensión de las capas ocultas
layers	Cantidad de capas ocultas
dropout	Probabilidad de apagar neuronas en cada capa
learning_rate	Tasa de aprendizaje
epoch	Número de épocas de entrenamiento
batch_size	Tamaño del batch para el entrenamiento
act	Función de activación en las capas ocultas
weight	Peso de la reconstrucción de la estructura de los nodos

2.5.1 Entrenamiento

En el entrenamiento con el modelo GAAN implicó una serie de decisiones críticas en relación con los hiperparámetros. Se probaron diferentes combinaciones de hiperparámetros, con un total de X pruebas realizadas para optimizar el rendimiento del modelo. La elección de estos es crucial, ya que tiene un impacto directo en la eficacia del modelo.

A continuación, se lista los valores de los diferentes hiperparámetros:

Tabla 2

Valores de los diferentes hiperparámetros

Hiperparámetros	Valor1	Valor2	Valor3	Valor4	Valor5
noise_dim	8	12	14	16	32
hid_dim	16	32	64	128	256
weight	0.2	0.3	0.4	0.5	0.7
epoch	100	200	300	400	500
layers	4	8	12	16	32
dropout	0.0	0.1	0.2	0.3	0.4

Durante este proceso de optimización, también se identificaron ciertas características de los datos que podrían necesitar ajustes, lo que nos llevó a (recurrencia) la etapa de análisis y preparación de datos en varias ocasiones. Esta conexión iterativa entre entrenamiento y análisis es vital para garantizar que el modelo este adecuadamente sintonizado y sea representativo de los datos.

2.5.1.1 Hiperparámetros de Iteraciones significativas

Tabla 3

Combinación de hiperparámetros más significativos en cada iteración

ITERACIONES	HID DIM	NOISE	LAYERS	DROPOUT	EPOCHS	WEIGHT	Nodos
1	256	16	16	0.4	500	0.7	1497
2	128	16	16	0.4	500	0.7	1497
3	64	16	16	0.3	400	0.5	1383
4	64	16	12	0.2	300	0.5	750
5	64	16	12	0.1	250	0.5	142

2.5.1.2 Hiperparámetros de Iteración final

Tabla 4

Hiperparámetros utilizados para la iteración final

HID	DIM	NOISE	LAYERS	DROPOUT	EPOCHS	WEIGHT	# Nodos
64		16	4	0.0	200	0.3	142

2.5.2 Validación

Dado que GAAN es un modelo de aprendizaje no supervisado, la validación no se basó en métricas tradicionales. En lugar de eso, se examinó el grafo generado para evaluar su calidad, prestando especial atención a la detección de nodos anómalos basados en sus características. Esta inspección cualitativa sirvió para identificar áreas de mejora y recalibrar el proceso de entrenamiento y preparación de datos.

2.5.2.1 Análisis de iteraciones

- **Primera iteración**

Validación: Al evaluar inicialmente el grafo generado, notamos que se otorgaba un énfasis excesivo a la estructura del grafo, y no a las características de los nodos. Esta observación nos condujo a regresar a la etapa de análisis y preparación de datos, donde optamos por normalizar los datos de todas las métricas. La normalización permitió que todas las características tuvieran el mismo rango, lo que ayudó a equilibrar la importancia entre la estructura del grafo y las propiedades de los nodos.

- **Segunda iteración**

Validación: En una revisión posterior, identificamos que ciertas métricas estaban fuertemente correlacionadas y no aportaban información significativa durante el entrenamiento. Al eliminar estas métricas, esperábamos mejorar la eficiencia del modelo y obtener un grafo más representativo.

- **Tercera iteración**

Validación: Otra inspección del grafo reveló nodos que presentaban 1 solo contrato ofertado y además su porcentaje de contratos únicos era 1. Decidimos filtrar estos nodos, ya que no aportaban información relevante en sus características y el número total de proveedores quedó en 1383.

- **Cuarta iteración**

Validación: Se procedió a verificar el grafo, y se encontraron nodos con un valor de grado de salida de 0 o 1, sesgando el rendimiento del modelo hacia aspectos estructurales. Decidimos filtrar estos nodos reduciendo el número total de 1383 a 750. Esta reducción tuvo como objetivo mejorar la generalización del modelo y reducir el sesgo introducido.

- **Quinta iteración**

Validación: A pesar de las mejoras realizadas, todavía observamos nodos que afectaban negativamente el modelo. En particular, aquellos nodos con pocos contratos. Al decidir conservar solo los nodos con al menos tres contratos colectivos, refinamos aún más nuestra representación a 142 nodos esenciales, lo que resultó en un grafo más limpio y representativo.

2.6 Construcción del grafo que modela las relaciones entre los proveedores

Tras finalizar las etapas de análisis y preparación de datos y entrenamiento y validación, procedimos a la construcción del grafo final. En este grafo, cada nodo del grafo correspondía a un proveedor y estaba caracterizado por un conjunto de atributos numéricos derivados de las métricas de transparencia definida en etapa 2.2. Por otro lado, cada arco representaba una relación entre dos proveedores.

A continuación, se presenta en la Tabla 5 la información sobre el conjunto de datos finales utilizados para la generación del grafo.

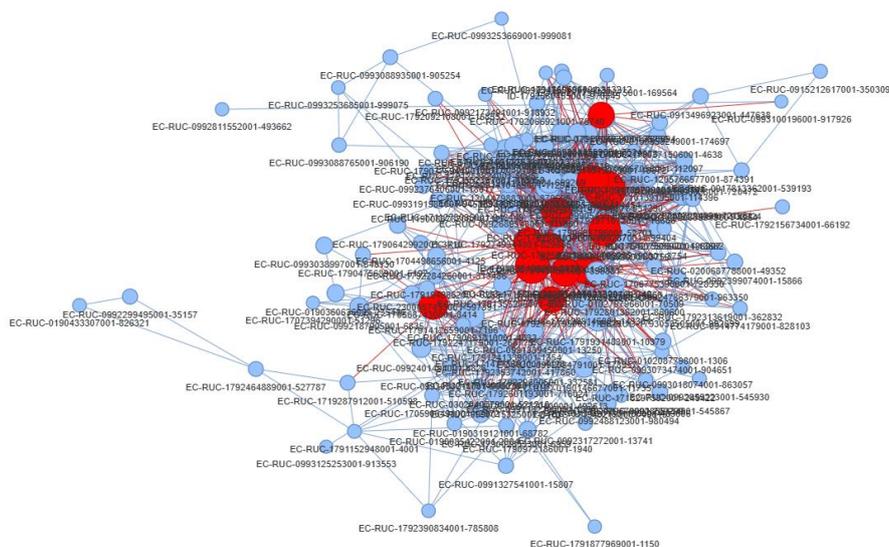
Tabla 5

Detalle sobre conjunto de datos finales utilizados en el análisis

Nombre	Detalle
Total de proveedores	142
Conexiones entre proveedores	1938
Características por proveedor	15

Figura 6

Red de Proveedores



Nota: Los nodos rojos son los proveedores que fueron identificados como anómalos por el modelo GAAN. La imagen es de autoría propia

2.7 Interpretación de los resultados

La interpretación de los resultados es una etapa crucial para comprender el impacto y la relevancia de nuestro modelo. Después de realizar varias iteraciones, ajustar los datos y el modelo, logramos construir un grafo más limpio y representativo de las relaciones entre los proveedores del sector de la salud.

2.7.1 Identificación de Proveedores Anómalos

El objetivo principal de nuestro estudio era caracterizar a los proveedores e identificar posibles nodos o proveedores anómalos en el sistema de compras públicas. Utilizando el modelo GAAN, pudimos identificar varios proveedores que presentaron características anómalas en

comparación con otros proveedores. Estos proveedores anómalos pueden ser objeto de un escrutinio más detenido para determinar si hay irregularidades o posibles fraudes en sus contratos.

2.7.2 Consideraciones éticas y legales

Es importante tener en cuenta las consideraciones éticas y legales al interpretar los resultados. Aunque el modelo identificó a varios proveedores como anómalos, esto no significa necesariamente que estén involucrados en actividades ilícitas o irregulares. Un proveedor puede ser identificado como anómalo debido a una variedad de razones, como un número inusualmente alto de contratos, una cantidad significativa de contratos únicos, o una estructura de relaciones inusuales. Por lo tanto, es fundamental no sacar conclusiones apresuradas sobre la legalidad o ética de un proveedor basándose únicamente en los resultados obtenidos.

Además, es crucial mantener la confidencialidad de los datos y respetar las leyes de protección de datos. Aunque los datos utilizados en este estudio son de dominio público, la identificación de proveedores anómalos puede tener implicaciones legales y reputacionales para los proveedores involucrados. Es por esto, que en el desglose de los resultados se optó por identificar a los proveedores anómalos con un ID numérico el cual respetaría la privacidad de cada individuo.

Capítulo 3

3. Resultados y análisis

3.1 Resultados del Modelo

De la totalidad de los proveedores analizados, el modelo identificó a 13 proveedores como anómalos. Este valor representa el 9.15% de los proveedores en el conjunto de datos. La identificación precisa de estas anomalías se basó en un enfoque integral que consideró tanto la estructura de las conexiones como las características individuales de los nodos. Esta combinación de factores permitió al modelo no solo resaltar posibles irregularidades en las relaciones entre proveedores, sino también discernir comportamientos inusuales a nivel individual.

3.1.1 Análisis de información estructural

En cuanto a la información estructural se tomó en cuenta el peso o tamaño de los nodos, que en el caso particular de este estudio se refiere al número total de contratos ofertados, y el grado de salida de cada nodo. Estos datos se muestran en la Tabla 6.

Tabla 6

Información de tipo estructural de los nodos anómalos identificados

ID	Contratos Ofertados	Grado de salida
1	67	32
2	63	43
3	52	34
4	53	35
5	84	53
6	105	37
7	52	27
8	64	21
9	63	28
10	64	25
11	74	36
12	56	34
13	75	36
Promedio	67	34

Se observa en la Tabla 6 que, entre todos los nodos anómalos, el promedio de contratos ofertados fue de 67. El número máximo de contratos ofertados fue de 105 para el nodo 6 y el número menor de contratos ofertados fue de 52 para el nodo 3 y el nodo 7. Generalmente los nodos

anómalos obtuvieron más de 50 contratos ofertados independientemente de si los ha ganado o perdido.

Entre todos los nodos anómalos el promedio de grado de salida (número de conexiones con otros nodos en el grafo) fue de 34, el grado de salida máximo fue 53 para el nodo 2 mientras que el menor grado de salida fue 21 para el nodo 8. Comúnmente los nodos anómalos tenían al menos 20 conexiones con otros nodos.

3.1.1.1 Comparativa con nodos no anómalos

Una comparativa efectiva entre nodos anómalos y no anómalos nos permitió entender las características distintivas entre los dos grupos y cómo estos factores se relacionaron con la anomalía observada.

Tabla 7

Información promedio de tipo estructural de los nodos no anómalos

	Contratos Ofertados	Grado de salida
Promedio	15	11

Se constató que, en promedio, los nodos anómalos ofertaron más de 4 veces la cantidad de contratos en comparación con los nodos no anómalos. Esta particularidad sugirió que los nodos anómalos eran, en general, proveedores más activos. Dada la alta cantidad de contratos ofertados, era plausible que estos nodos actuaran como rellenos en ocasiones.

Por otro lado, los nodos anómalos, en promedio, poseían más de 3 veces el grado de salida comparado con los nodos no anómalos. Esta amplia interconexión señalaba una mayor interacción entre ellos, sugiriendo un nivel de colaboración o relación más estrecha en su ámbito de operación.

3.1.2 Análisis de métricas para contratos perdidos

A continuación, en la Tabla 8, se muestran los valores de las características o métricas que están involucradas con el comportamiento del proveedor en los contratos que ha perdido, tal como se explicó en la sección *Métricas contratos perdidos*.

Tabla 8

Desglose de métricas para los contratos perdidos de cada nodo anómalo

ID	PCP	DEPCP	PDUOMGCP	PDUOMPCP	PFCP	POCP	PICCP	RICCP
1	0,582	0,108	0,167	0,250	0,270	0,837	0,240	0,959
2	0,714	0,076	0,322	0,171	0,208	0,888	0,200	0,419
3	0,519	0,084	0,346	0,208	0,208	0,882	0,188	0,436
4	0,736	0,117	0,249	0,250	0,200	0,840	0,249	0,628
5	0,798	0,104	0,220	0,162	0,258	0,902	0,209	0,689
6	0,419	0,131	0,249	0,309	0,250	0,799	0,185	0,380
7	0,615	0,109	0,202	0,220	0,329	0,857	0,235	0,857
8	0,656	0,076	0,344	0,161	0,215	0,893	0,200	0,439
9	0,619	0,095	0,224	0,265	0,233	0,847	0,263	0,477
10	0,500	0,109	0,206	0,266	0,231	0,831	0,259	0,535
11	0,676	0,126	0,171	0,234	0,325	0,876	0,196	0,426
12	0,679	0,123	0,322	0,284	0,256	0,818	0,215	0,368
13	0,840	0,063	0,285	0,139	0,190	0,899	0,209	0,556
Promedio	0,643	0,102	0,255	0,224	0,244	0,859	0,219	0,552

A continuación, se presenta una explicación más detallada del desglose de estas métricas:

- Entre todos los nodos anómalos el PCP promedio fue de un 64%, el valor mínimo fue 41% para el nodo 6 y el valor máximo fue 84% para el nodo 13. De manera general, los nodos tuvieron un PCP de al menos 40%.
- Entre todos los nodos anómalos la DEPCP promedio fue de 0,102; el valor mínimo fue 0,063 para el nodo 13 y el valor máximo fue 0,126 para el nodo 11. De manera general, los nodos tuvieron una DEPCP de al menos 0,063.
- Entre todos los nodos anómalos el PDUOMGCP promedio fue de 0,255; el valor mínimo fue 0,167 para el nodo 1 y el valor máximo fue 0,346 para el nodo 3. De manera general, los nodos tuvieron un PDUOMGCP de al menos 0,167.
- Entre todos los nodos anómalos la PDUOMPCP promedio fue de 0,224; el valor mínimo fue 0,139 para el nodo 13 y el valor máximo fue 0,309 para el nodo 6. De manera general, los nodos tuvieron una PDUOMPCP de al menos 0,139.
- Entre todos los nodos anómalos la PFCP promedio fue de 0,244; el valor mínimo fue 0,190 para el nodo 13 y el valor máximo fue 0,329 para el nodo 7. De manera general, los nodos tuvieron una PFCP de al menos 0,190.

- Entre todos los nodos anómalos el POCP promedio fue de un 0,859; el valor mínimo fue 0,799 para el nodo 6 y el valor máximo fue 0,902 para el nodo 5. De manera general, los nodos tuvieron un POCP de al menos 0,799.
- Entre todos los nodos anómalos el PICCP promedio fue de 0,219; el valor mínimo fue 0,185 para el nodo 6 y el valor máximo fue 0,263 para el nodo 9. De manera general, los nodos tuvieron un PICCP de al menos 0,185.
- Entre todos los nodos anómalos el RICCP promedio fue de 0,552; el valor mínimo fue 0,368 para el nodo 12 y el valor máximo fue 0,959 para el nodo 1. De manera general, los nodos tuvieron un RICCP de al menos 0,368.

3.1.2.1 Comparativa con nodos no anómalos

Tabla 9

Desglose promedio de métricas para los contratos perdidos de los nodos no anómalos

ID	PCP	DEPCP	PDUOMGCP	PDUOMPCP	PFCP	POCP	PICCP	RICCP
Promedio	0.527	0.084	0.219	0.179	0.229	0.829	0.219	0.290

Dentro del análisis realizado de las métricas, ciertas observaciones se alinearon con la noción de que algunos nodos anómalos actuaron como rellenos o que podrían haber estado involucrados en prácticas de colusión. Profundizamos en estos hallazgos:

3.1.2.1.1 Porcentaje de Contratos Perdidos PCP

Nodos Anómalos: Promedio de 64.3%.

Nodos No Anómalos: Promedio de 52.7%.

El notablemente alto PCP en nodos anómalos, que supera en más del 11% a los no anómalos, puede indicar que estos nodos presentan ofertas sin ninguna genuina intención de ganar. Si conscientemente hicieron ofertas, pero no lograron ganar contratos, fue un indicativo de que estuvieron allí para aumentar la cantidad de oferentes, actuando como relleno y dando la apariencia de competencia.

3.1.2.1.2 Desviación Estándar Promedio de los Contratos Perdidos (DEPCP)

Nodos Anómalos: Promedio de 0.102.

Nodos No Anómalos: Promedio de 0.084.

Con estos valores, se revelo que hubo una mayor variabilidad en las ofertas acumuladas de los nodos anómalos. Esto indica que, debido a la gran cantidad de contratos que han perdido en comparación a los nodos no anómalos, las ofertas de los nodos anómalos muestran una diferencia mayor entre sí. Esto apoya la noción de un comportamiento de relleno.

3.1.2.1.3 Promedio de Diferencia de la Última Oferta con Respecto al Monto Ganador en Contratos Perdidos (PDUOMGCP) y Promedio de Diferencia de la Última Oferta con Respecto al Monto Planeado en Contratos Perdidos (PDUOMPCP):

Tabla 10

Desglose promedio de las métricas PDUOMGCP y PDUOMPCP para nodos anómalos y no anómalos

	PDUOMGCP	PDUOMPCP
Anómalos	0.255	0.224
No Anómalos	0.219	0.179

Los nodos anómalos tendieron a ofrecer montos que diferían en mayor medida con respecto al monto referencial en comparación a los nodos no anómalos. Dado que los nodos anómalos tienen un grado de salida mucho mayor que los no anómalos, esto implica que en sus contratos hubo más proveedores con los cuales se debía competir. Se podría decir que una mayor competencia obliga a que el proveedor oferte montos más alejados al monto referencial, incluso si está como actor de relleno.

3.1.2.1.4 Promedio de Ofertas Perdidas (POCP)

Nodos Anómalos: Promedio de 0.859

Nodos No Anómalos: Promedio de 0.829

El hecho de que los nodos anómalos perdieran una proporción significativamente mayor de sus ofertas, más allá de la diferencia porcentual, sugirió que su objetivo principal podría no haber sido ganar la subasta, sino influir en ella de alguna manera. Esta influencia pudo haber sido

para ayudar a otro proveedor, tal vez uno con el que estuvieron coludidos, a tener una mejor oportunidad de ganar.

3.1.2.1.5 Promedio de Frecuencia de Contrato Colectivos Perdidos (PFCP)

Nodos Anómalos: Promedio de 0.244.

Nodos No Anómalos: Promedio de 0.229.

El PFCP fue ligeramente mayor en los nodos anómalos. Si bien la diferencia no fue drásticamente alta, aun insinúa que los nodos anómalos tendieron a ofertar poco en los contratos que perdieron, lo que pudo haber sido otra indicación de su papel como relleno.

3.1.2.1.6 Promedio Índice de Clasificación Contratos Perdidos (PICCP):

Nodos anómalos: 0.219

Nodos no anómalos: 0.219

Los nodos anómalos tienen un promedio de índice de clasificación igual al de los no anómalos. Esto indica que en los contratos que ganaron, la mayoría de los proveedores pasaron la etapa de clasificación para la subasta.

Tomando en cuenta los análisis previos, se concluye que en estos casos el proveedor anómalo ha coludido con otros proveedores en sus contratos para tomar el rol de relleno y permitan a esos proveedores ganar la subasta.

3.1.2.1.7 Rango Índice de Clasificación Contratos Perdidos (RICCP):

Nodos anómalos: 0.552

Nodos no anómalos: 0.290

Los nodos anómalos tienen un rango de índice de clasificación mayor al de los no anómalos. Esto indica que tienen una variedad de valores más alta que los nodos no anómalos, indicando que su comportamiento es menos consistente en cuanto a sus contratos perdidos.

Tomando en cuenta los análisis previos, se concluye que en estos casos el proveedor anómalo ha seguido indicaciones de participar en contratos donde el número de proveedores que

pasan la etapa de clasificación varía considerablemente. Esto apoya la noción de tomar un rol de relleno en otros contratos.

3.1.3 *Análisis de métricas para contratos ganados*

A continuación, en la Tabla 11, se muestran los valores de las características o métricas que están involucradas con el comportamiento del proveedor en los contratos en los que ha ganado la subasta inversa, tal como se explicó en la sección *Métricas contratos ganados*.

Tabla 11

Desglose de métricas para los contratos ganados de participación colectiva de cada nodo anómalo

ID	PCC	DEPCC	PDUOCC	POCC	PFCC	PMGCC	PICCC	RICCC
1	0,418	0,127	0,247	0,876	0,608	0,062	0,252	0,477
2	0,286	0,118	0,304	0,854	0,457	0,082	0,234	0,493
3	0,481	0,115	0,272	0,868	0,457	0,024	0,210	0,641
4	0,264	0,112	0,246	0,855	0,432	0,048	0,208	0,465
5	0,202	0,170	0,261	0,868	0,517	0,060	0,279	0,612
6	0,581	0,145	0,309	0,846	0,508	0,063	0,233	0,720
7	0,385	0,192	0,394	0,790	0,565	0,091	0,279	0,505
8	0,344	0,101	0,192	0,922	0,670	0,108	0,229	0,301
9	0,381	0,081	0,201	0,888	0,409	0,022	0,286	0,602
10	0,500	0,110	0,262	0,826	0,463	0,077	0,358	0,505
11	0,324	0,155	0,282	0,851	0,638	0,033	0,274	0,505
12	0,321	0,167	0,385	0,791	0,519	0,089	0,269	0,558
13	0,160	0,086	0,171	0,919	0,502	0,064	0,254	0,380
Promedio	0,357	0,129	0,271	0,858	0,519	0,063	0,259	0,520

A continuación, se presenta una explicación más detallada del desglose de estas métricas:

- Entre todos los nodos anómalos el PCC varía entre 0.160 y 0.581 con un promedio de 0.357. Esto indica una variabilidad en el porcentaje de contratos colectivos ganados por los diferentes proveedores. Mientras que el nodo 13 tiene el valor más bajo, indicando una tasa de éxito baja a la hora de ganar contratos, el nodo 6 presenta el valor más alto, lo que sugiere un desempeño superior al resto de nodos.
- Entre todos los nodos anómalos la DEPCC promedio fue 0,129; el valor mínimo fue de 0,086 para el nodo 13, mientras que el valor máximo fue de 0,192 para el proveedor 7.

- Entre todos los nodos anómalos el PDUOCC promedio fue de un 0,271; el valor mínimo fue 0,171 para el nodo 13 y el valor máximo fue 0,394 para el nodo 7.
- Entre todos los nodos anómalos el POCC promedio fue de un 0,858; el valor mínimo fue 0,790 para el nodo 7 y el valor máximo fue 0,922 para el nodo 8. De manera general, los nodos tuvieron un POCC de al menos 0,790.
- Entre todos los nodos anómalos la PFCC promedio fue de 0,519; el valor mínimo fue 0,409 para el nodo 9 y el valor máximo fue 0,670 para el nodo 8. De manera general, los nodos tuvieron una PFCC de al menos 0,40.
- Entre todos los nodos anómalos el PMGCC promedio fue de 0.063; el valor mínimo de 0.022 para el nodo 9 y el valor máximo de 0.108 para el nodo 8.
- Entre todos los nodos anómalos el PICCC promedio fue de 0,259; el valor mínimo fue 0,208 para el nodo 4 y el valor máximo fue 0,358 para el nodo 10. De manera general, los nodos tuvieron un PICCP de al menos 0,20.
- Entre todos los nodos anómalos el RICCC promedio fue de 0,520; el valor mínimo fue 0,301 para el nodo 8 y el valor máximo fue 0,720 para el nodo 6. De manera general, los nodos tuvieron un RICCC de al menos 0,30.

3.1.3.1 Comparativa con nodos no anómalos

Tabla 12

Desglose promedio de métricas para los contratos ganados de participación colectiva nodos no anómalos

	PCC	DEPCC	PDUOCC	POCC	PFCC	PMGCC	PICCC	RICCC
Promedio	0,473	0,108	0,231	0,885	0,490	0,103	0,284	0,310

3.1.3.1.1 Porcentaje de Contratos Colectivos por Proveedor (PCC)

Nodos anómalos: 0.357

Nodos no anómalos: 0.473

El porcentaje de contratos colectivos de los nodos anómalos fue menor en comparación con los nodos no anómalos. Esta diferencia insinuó que, aunque estuvieran participando, los nodos anómalos no tuvieron la misma tasa de éxito que los no anómalos. Este comportamiento podría haber sido un indicador de que actuaran como relleno, estando presentes más para dar una ilusión de competencia que para ganar contratos.

3.1.3.1.2 Desviación Estándar Promedio de los Contratos Colectivos (DEPCC):

Nodos anómalos: 0.129

Nodos no anómalos: 0.108

El valor para los nodos anómalos fue ligeramente mayor en los nodos anómalos, sugiriendo una variabilidad mayor en sus ofertas ganadoras en comparación con los no anómalos. Esta variabilidad podría indicar una posible colusión, donde ciertos proveedores se ajustaban estratégicamente a sus ofertas para favorecerlo.

3.1.3.1.3 Promedio de la Diferencia de la Última Oferta con respecto al Monto Planeado en Contratos Colectivos (PDUOCC):

Nodos anómalos: 0.271

Nodos no anómalos: 0.231

Los nodos anómalos mostraron una mayor diferencia entre su última oferta y el monto planeado, lo que podía haber sido un indicio de intentos de aumentar o disminuir los precios de las ofertas.

3.1.3.1.4 Promedio de ofertas colectivas (POCC):

Nodos anómalos: 0.858

Nodos no anómalos: 0.885

La similitud en el promedio de ofertas colectivos entre ambos grupos no mostró diferencias notables. Sin embargo, las demás métricas ofrecieron una imagen más completa del comportamiento de los nodos anómalos.

3.1.3.1.5 Promedio de Frecuencia de Contratos Colectivos Ganados (PFCC):

Nodos anómalos: 0.519

Nodos no anómalos: 0.490

Aunque los nodos anómalos ganaron contratos colectivos con una frecuencia ligeramente mayor, esto no fue una garantía de un comportamiento normal. Combinado con otras métricas, este ligero incremento podía haber sido una táctica para desviar la atención de otras prácticas irregulares.

3.1.3.1.6 Promedio Monto Ganado en Contratos Colectivos (PMGCC):

Nodos anómalos: 0.063

Nodos no anómalos: 0.103

El monto promedio ganado por nodos anómalos fue considerablemente menor. Esta diferencia significativa pudo haber sido una indicación de que estaban dispuestos a ofertar a precios muchos más bajos.

3.1.3.1.7 Promedio Índice de Clasificación Contratos Colectivos (PICCC):

Nodos anómalos: 0.259

Nodos no anómalos: 0.284

Los nodos anómalos tienen un promedio de índice de clasificación menor al de los no anómalos. Esto indica que en los contratos que ganaron, la mayoría de los proveedores pasaron la etapa de clasificación para la subasta. Tomando en cuenta los análisis previos, se concluye que en estos casos el proveedor anómalo ha coludido con otros proveedores en sus contratos para que hagan el rol de relleno y le permitan ganar la subasta.

3.1.3.1.8 Rango Índice de Clasificación Contratos Colectivos (RICCC):

Nodos anómalos: 0.520

Nodos no anómalos: 0.310

Los nodos anómalos tienen un rango de índice de clasificación mayor al de los no anómalos. Esto indica que tienen una variedad de valores más alta que los nodos no anómalos, indicando que su comportamiento es menos consistente en cuanto a sus contratos ganados. Tomando en cuenta los análisis previos, se concluye que en estos casos el proveedor anómalo ha seguido indicaciones de participar en contratos donde el número de proveedores que pasan la etapa de clasificación varía considerablemente. Esto apoya la noción de usar tácticas para cubrir su rol de relleno en otros contratos.

3.1.4 Análisis de caso excepción – proveedores de participación única

En la sección Generación de Características de Transparencia se explicaron los hallazgos de un caso particular de comportamiento en los proveedores: los proveedores de participación única. Estos proveedores fueron removidos de los datos finales debido a que causaban ruido en el modelo; sin embargo, se obtuvieron datos significativos que son importantes de abordar, aunque sea de manera superficial.

En total se obtuvieron 136 proveedores de participación única. Los promedios de sus métricas más significativas se muestran en la Tabla 13.

Tabla 13

Desglose de métricas para los contratos ganados de proveedores de participación única

	PCU	Ofertados	PICCU	RICCU
Promedio	1	1	0,453	0,011

A continuación, se explican los valores mostrados en la tabla:

- En promedio, los proveedores de participación única ofertaron 1 contrato en el año 2020.
- En promedio su PICCU fue de 0,453; un valor que puede considerarse relativamente alto.
- En promedio su RICCU fue de 0,011; un valor sumamente bajo lo cual indica que la mayoría de estos proveedores tuvo un RICCU de 0, evidenciado por el promedio de contratos ofertados siendo de 1.

Capítulo 4

4. Conclusiones y recomendaciones

4.1 Conclusiones

- Se extrajeron características significativas de los proveedores de bienes y servicios en base a la información de los procesos de adquisiciones públicas que facilitó su descripción detallada.
- Se desarrolló un modelo que identificó proveedores anómalos de bienes y servicios de un sistema de compras públicas, utilizando técnicas de aprendizaje automático.
- Se crearon visualizaciones de los proveedores anómalos que participan en el proceso de contratación que facilitó su análisis a organismos de control.
- Es posible identificar patrones de comportamiento para los proveedores de bienes y servicios de un sistema de compras públicas en base a la caracterización significativa de estos.

4.2 Recomendaciones

- Un aspecto por abordar es la inclusión de las entidades contratantes como nodos en el grafo de entrada. Esto podría aportar con un análisis mucho más robusto ya que sería una caracterización tanto de los proveedores que participan en contratos como de las entidades que los publican. Identificar no solo proveedores anómalos sino también entidades anómalas, podría resultar en hallazgos como redes de colaboración o colusión entre proveedores y entidades.
- Los proveedores que poseen participación única tienen información pertinente que podría analizarse en un futuro de manera aislada. Este tipo de contratos suceden en situaciones de emergencia y no reflejan el comportamiento usual, pero como caso aislado se consideran importantes los hallazgos que se pueden obtener de su análisis.
- Considerar analizar diferentes representaciones de grafos para los proveedores, o, los contratos. ¿Qué tal si el grafo fuese bipartito? ¿O unidireccional en vez de bidireccional?

Existe mucho potencial de análisis al cambiar ciertas características estructurales del grafo, y, al escoger la técnica de aprendizaje automático más adecuada para cada tipo de grafo.

5. Referencias

- [1] B. Soto y V. Toledo, «Catálogo Bibliográfico General PUCV,» Marzo 2017. [En línea]. Available: http://opac.pucv.cl/pucv_txt/txt-8500/UCD8532_01.pdf. [Último acceso: 26 Agosto 2023].
- [2] «Universidad de Granada,» [En línea]. Available: <https://ccia.ugr.es/~jfv/ed1/c++/cdrom4/paginaWeb/grafos.htm>.
- [3] J. Frost, «statisticsbyjim,» Abril 2018. [En línea]. Available: <https://statisticsbyjim.com/basics/normal-distribution/>. [Último acceso: 26 Agosto 2023].
- [4] W. Badr, «towardsdatascience,» 22 Abril 2019. [En línea]. Available: <https://towardsdatascience.com/auto-encoder-what-is-it-and-what-is-it-used-for-part-1-3e5c6f017726>. [Último acceso: 26 Agosto 2023].
- [5] J. Jiang, «towardsdatascience,» 5 Mayo 2021. [En línea]. Available: <https://towardsdatascience.com/what-is-embedding-and-what-can-you-do-with-it-61ba7c05efd8>. [Último acceso: 26 Agosto 2023].
- [6] T. Kipf y M. Welling, «Semi-Supervised Classification with Graph Convolutional Networks,» *CoRR*, vol. abs/1609.02907, 2016.
- [7] K. Yasar, «TechTarget,» Marzo 2023. [En línea]. Available: <https://www.techtarget.com/searchenterpriseai/definition/generative-adversarial-network-GAN>. [Último acceso: 2026 Agosto 2023].
- [8] K. Ding, J. Li, R. Bhanushali y H. Liu, «Deep Anomaly Detection on Attributed Networks,» de *SIAM International Conference on Data Mining (SDM)*, 2019.
- [9] Z. Chen, B. Liu, M. Wang, P. Dai, J. Lv y L. Bo, «Generative Adversarial Attributed Network Anomaly Detection,» de *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, Association for Computing Machinery, 2020, p. 1989–1992.
- [10] K. Liu, Y. Dou, Y. Zhao, X. Ding, X. Hu, R. Zhang, K. Ding, C. Chen, H. Peng, K. Shu, G. H. Chen, Z. Jia y P. S. Yu, «PyGOD: A Python Library for Graph Outlier Detection,» *arXiv preprint arXiv:2204.12095*, 2022.
- [11] M. Fey y J. E. Lenssen, «Fast Graph Representation Learning with {PyTorch Geometric},» de *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

- [12] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga y A. Lerer, «PyTorch: An Imperative Style, High-Performance Deep Learning Library,» de *Advances in Neural Information Processing Systems 32*, Curran Associates, Inc., 2019, pp. 8024--8035.
- [13] R. Correa, «Creación, Administración y Desarrollo del portal del Sistema Oficial de Información de Contratación Pública y Consultoría del Ecuador,» 14 Noviembre 2007. [En línea]. Available: https://portal.compraspublicas.gob.ec/sercop/wp-content/uploads/files/100/Decreto_744.pdf. [Último acceso: 31 Mayo 2023].
- [14] X. Wang, B. Jin y Y. Du, «One-class graph neural networks for anomaly detection in attributed networks,» *Neural Comput & Applic* , p. 12073–12085, 2021.
- [15] M. E. K. Niessen, J. M. Paciello y J. I. P. Fernandez, «Anomaly Detection in Public Procurements using the Open Contracting Data Standard,» *2020 Seventh International Conference on eDemocracy & eGovernment (ICEDEG)*, pp. 127-134, 2020.
- [16] X. Ma, J. Wu, S. Xue, J. Yang, C. Zhou, Q. Z. Sheng, H. Xiong y L. Akoglu, «A Comprehensive Survey on Graph Anomaly Detection with Deep Learning,» *IEEE Transactions on Knowledge and Data Engineering*, pp. 1-1, 2021.
- [17] J. Alonso, «Incentro,» 15 Junio 2022. [En línea]. Available: <https://www.incentro.com/es-ES/blog/que-es-tensorflow>. [Último acceso: 5 Diciembre 2022].
- [18] «Open Contracting Data Standard — Open Contracting Data Standard 1.1.5 documentation,» [En línea]. Available: <https://standard.open-contracting.org/latest/es/primer/what/>. [Último acceso: 1 Junio 2023].
- [19] «IONOS Digital Guide,» [En línea]. Available: <https://www.ionos.es/digitalguide/online-marketing/marketing-para-motores-de-busqueda/graph-neural-network/>. [Último acceso: 1 Junio 2023].
- [20] «Ionos,» 8 Octubre 2020. [En línea]. Available: <https://www.ionos.es/digitalguide/online-marketing/marketing-para-motores-de-busqueda/que-es-keras/>. [Último acceso: 5 Diciembre 2022].
- [21] «Contrataciones Abiertas Ecuador - OCDS,» [En línea]. Available: <https://datosabiertos.compraspublicas.gob.ec/PLATAFORMA/acerca>. [Último acceso: 1 Junio 2023].
- [22] «Compraspublicas,» [En línea]. Available: <https://portal.compraspublicas.gob.ec/sercop/la-institucion/>. [Último acceso: 6 Junio 2023].

- [23] «Blog oficial de NVIDIA Latino América,» [En línea]. Available: <https://la.blogs.nvidia.com/2022/10/31/que-son-las-graph-neural-networks/>. [Último acceso: 1 Junio 2023].
- [24] «Aprendeconalf,» 05 Mayo 2022. [En línea]. Available: <https://aprendeconalf.es/docencia/python/manual/numpy/>. [Último acceso: 5 Diciembre 2022].
- [25] K. Liu, Y. Dou, Y. Zhao, X. Ding, X. Hu, R. Zhang, K. Ding, C. Chen, H. Peng, K. Shu, L. Sun, J. Li, G. H. Chen, Z. Jia y P. S. Yu, «Bond: Benchmarking unsupervised outlier node detection on static attributed graphs,» *Advances in Neural Information Processing Systems*, vol. 35, pp. 27021--27035, 2022.