PREDICCIÓN DEL RIESGO DE CORRUPCIÓN EN LOS PROCESOS DE CONTRATACIÓN PÚBLICA DISPONIBLES EN LA PLATAFORMA DATOS ABIERTOS USANDO EL MODELO GRADIENT BOOSTING MACHINE

Corruption risk prediction in public procurement processes available on Datos Abiertos platform using Gradient Boosting Machine model

María Fernanda Umbarila Suárez[1]

[1]Universidad Libre, Bogotá, Colombia, *mariaf-umbarilas@unilibre.edu.co*

Fecha de recepción: 30/08/2024

RESUMEN

La corrupción y los mecanismos para combatirla se han vuelto un tema de alta importancia para los gobiernos alrededor del mundo, Colombia junto a otros países de la OCDE cuenta con bases de datos abiertas y sistemas de contratación pública que son actualizadas constantemente y las cuales están conformadas por una cantidad grande de datos, lo cual plantea un reto para los organismos de control que se enfrentan a la tarea de auditar y analizar dicha información. El Machine Learning se presenta entonces como una herramienta para facilitar la auditoría de contratación pública, al ofrecer un mecanismo rápido para identificar aquellos procesos contractuales con un riesgo mayor de incidir en corrupción. Así, el presente artículo tiene como objetivo evaluar el rendimiento de un modelo de Gradient Boosting Machine al momento de predecir el riesgo de corrupción en procesos de contratación pública disponible en la plataforma Datos Abiertos.

Palabras clave: Contratación pública, corrupción, plataforma Datos Abiertos, predicción, Gradient Boosting Machine.

ABSTRACT

Corruption and the mechanisms to combat it have become an issue of high importance for governments around the world, Colombia along with other OECD countries has open databases and public procurement systems that are constantly updated, and which are made up of a large amount of data, situation that represent a challenge for

the control agencies faced with the task of auditing and analyzing such information. Machine Learning is then presented as a tool to facilitate the audit of public procurement, by offering a quick mechanism to identify those contractual processes with a higher risk of corruption. Thus, this article aims to evaluate the performance of a Gradient Boosting Machine model when predicting the risk of corruption in public procurement processes available on the Datos Abiertos platform.

Keywords: Corruption, Datos Abiertos platform, Gradient Boosting Machine, prediction, public procurement.

1. INTRODUCCIÓN

En Colombia se han venido desarrollando varias estrategias a nivel Estado para que los índices de corrupción sean menores, los cuales incluyen organismos como Transparencia por Colombia, el Programa de Transparencia y Ética Pública, los observatorios y estrategias de la Auditoría Nacional de Colombia, Contralorías y demás entidades de vigilancia, así como el proceso de implementación de ODCS que se ha enfocado en unificar las plataformas existentes y ofrecer datos abiertos que permitan realizar seguimientos eficaces a la contratación pública [1].

Dada la importancia que la detección y prevención de la corrupción tiene, los organismos de control gubernamentales y ciudadanos se enfrentan a la tarea de analizar la información de contratación de todo el país, a lo que la información de contratación publicada en Datos Abiertos se presenta como la principal fuente de información de consulta. Este portal contiene un set de datos con un total de 59 columnas sobre los contratos celebrados en SECOP II a nivel nacional y que, para mediados de agosto de 2024 cuenta con 5.9 millones de registros, lo que plantea un reto al momento de realizar un análisis o auditoría en búsqueda de anomalías, patrones e inconsistencias que puedan interpretarse como banderas rojas, y que puede retrasar el reporte de procesos de contratación sospechosos por parte de las entidades y ciudadanos.

Teniendo en cuenta los mecanismos de transparencia y acceso a datos abiertos dictados por las entidades mencionadas, y bajo los casos de éxito y el potencial mostrado que el Machine Learning tiene para la predicción de indicadores de riesgo de corrupción, este trabajo tiene como objetivo evaluar el rendimiento de un modelo de Gradient Boosting Machine al momento de predecir el riesgo de corrupción en procesos de contratación pública disponible en la plataforma Datos Abiertos.

El artículo se estructura en una revisión bibliográfica, donde se exploran los trabajos sobre los que se sustenta la investigación y cómo se busca complementarlos, se continúa con una descripción del conjunto de datos en la sección de área de estudio y metodología, se describe el modelo de Machine Learning a usar en la sección modelos, sigue con la descripción de los criterios de evaluación del modelo en la sección de criterios de rendimiento, por último, se continúa con la implementación del modelo y la interpretación de sus métricas de

ISSN 2322-8415 / Año 2024 / No. 24 - Vol. 14 / Páginas 61-70 Universidad Libre - Facultad de Ingeniería

rendimiento en las secciones de resultados y discusiones y conclusiones.

2. REVISIÓN BIBLIOGRÁFICA

Se puede entender la corrupción como el abuso de posiciones de poder para el beneficio particular de actores legales y/o ilegales, o en posible asociación entre estos, en detrimento del interés colectivo. Un informe de la Radiografía de hechos de corrupción en Colombia se destaca al identificar 967 hechos de corrupción reportados entre 2016 y 2020; este informe enfatiza que el ámbito contratación fue en el que se reportó mayor incidencia, con un 42% dentro del tipo administrativo, estos incluyen prácticas como el direccionamiento de contratos con requisitos habilitantes muy específicos, el pago del bien o servicio a pesar de no haberse cumplido en su totalidad, con lo cual se puede evidenciar que las irregularidades se presentan en las diferentes etapas de los procesos contractuales [2].

En este panorama, la prevención y detección de la ineficiencia, el fraude y la prevaricación puede ser tan importante como las sanciones que se impongan posteriormente [3]. La metodología de compilación de banderas rojas o alertas de corrupción en los procesos de contratación pública a partir de bases de datos públicas y estandarizadas se presenta entonces como uno de los primeros pasos al implementar un enfoque de control preventivo; una vez definidas dichas banderas, se prosigue a definir indicadores: características cuantitativas en los procesos de contratación que permitirán diferenciar unos procesos de contratación de otros, la disponibilidad de información es crucial para desarrollar buenos indicadores de corrupción, debe tenerse en cuenta que estos indicadores pueden ser manipulados para evitar la detección de irregularidades y que lo que se considera corrupción puede cambiar con el tiempo debido a la normativa local y actual, lo que implica que las métricas deben adaptarse a estos cambios normativos y al diseño institucional de las agencias de contratación [4].

Al usar banderas rojas e indicadores para la detección del riesgo de corrupción es importante considerar la calidad y estructura de los datos que vayan a analizarse. En el caso de Colombia, un diagnóstico realizado por la OEA [5] destaca que el país cuenta con tres plataformas en el Sistema Electrónico para la Contratación Pública (SECOP): SECOP I, SECOP II, además de la Tienda Virtual del Estado Colombia (TVEC). Datos Abiertos cuenta desde el 2014 con la información de los procesos de contratación registrados en SECOP I y II. Este mismo diagnóstico señala que en SECOP I hay múltiples errores en los valores de los contratos que, de forma individual, no son muy grandes, pero agregados representan un sesgo en la muestra de datos de varios billones de pesos; mientras que en SECOP II se encuentran errores de registro en pocos contratos, pero sesgan la muestra por un valor muy alto.

Así mismo, una autoría de datos realizada por Contreras Ussa [1] a los datos de contratación en Colombia concluyó que se pueden encontrar variedad de errores en los set de datos, que van desde digitación incorrecta, campos sin diligenciar, fechas de inicio y final que no son congruentes, valores de contratación cero, valores de contratos con inexactitudes que distan del precio real a contratar, entre otros, a lo cual los autores sugieren usar modelos de aprendizaje no supervisado para la corrección de datos y lograr así un mayor ajuste a modelos de datos abiertos como los de la OCDE.

Otra metodología para la detección de corrupción cuando no se cuenta con bases de datos públicas o estandarizadas se basa en estadísticas generales a lo largo de los años del estudio sobre compradores, proveedores, tipos de contratación, usando el monto y la cantidad de adquisiciones.

Con estos indicadores pueden desarrollarse modelos como el de Myháły Fazekas, el cual se alimenta de las bases de datos de registros de contrataciones públicas, aportes a campañas electorales, listados de proveedores entre otros, y que permite detectar patrones o banderas rojas sospechosas en contratos realizados por diferentes entidades [4].

Sin embargo, ambos enfoques pueden combinarse para lograr resultados con un mayor alcance, como es el caso del trabajo realizado por Gallego et al. [3], en el cual se usaron dos fuentes de información para lograr la predicción del riesgo de corrupción: la disponible en Datos Abiertos y la dispuesta por organismos de control como la Contraloría y la Confecámaras.

En un artículo de revisión, Modrušan et al. [6] analiza los diferentes modelos que han sido implementados para dicho fin, así como las métricas medidas en cada caso. El análisis pudo observar que se han utilizado dos enfoques principales: el aprendizaje supervisado y el no supervisado.

Estos métodos difieren en las variables objetivo, en el aprendizaje supervisado las variables objetivo se definen con precisión como salida del modelo, mientras que en el aprendizaje no supervisado no se tienen dichas variables preestablecidas. En este mismo estudio se señalan los modelos y variables objetivos usados según lo que se quiera detectar (casos concretos de corrupción, patrones, clústeres de información, etc.).

El trabajo desarrollado por Gallego et al. [3] permite evaluar las ventajas del uso de modelos de Machine Learning al momento de predecir el riesgo de corrupción, sin embargo, también concluyen que se presentan retos como el sesgo que pueden presentar las listas negras de la Contraloría y Confecámaras.

Adicional a esto, el análisis y entrenamiento de los modelos de Machine Learning fueron realizados sobre contratos celebrados entre el 2011 y el 2015.

Así, el presente trabajo se sustenta en evaluar el rendimiento de un modelo de Gradient Boosting Machine alimentado con los datos de listas negras, sanciones y adiciones extraídas de Datos Abiertos, los cuales son actualizados diariamente y de forma independiente a los contratistas y entidades contratantes, además, se tendrán en cuenta los procesos de contratación celebrados entre el 2015 y agosto del 2024, ampliando el alcance en tiempo del análisis.

3. ÁREA DE ESTUDIO Y METODOLOGÍA

3.1. Descripción del conjunto de datos

La plataforma Datos Abiertos cuenta con diferentes sets de datos relacionados a la contratación realizada en SECOP I, SECOP II, y sus metadatos asociados, todos alimentados por el administrador de estos sistemas: Colombia Compra Eficiente. En el presente trabajo se usarán tres sets específicos de datos, los cuales se describen en la Tabla 1.

ISSN 2322-8415 / Año 2024 / No. 24 - Vol. 14 / Páginas 61-70 Universidad Libre - Facultad de Ingeniería

Tabla 1.
Descripción de los sets de datos.

Nombre	Cantidad de columnas	Cantidad aproximada de filas
SECOP II - Procesos de Contratación	59	5.97 millones
SECOP II -Adiciones	5	4.72 millones
SECOPII - Multas y Sanciones	17	936

Debido a los problemas de calidad de datos encontrados por la OEA [5] y por Contreras Ussa [1] se limitó el alcance de los contratos a analizar con la intención de trabajar con datos lo más limpios posibles, realizando consultas a la API de Datos Abiertos con los siguientes condicionales principales:

- Departamento: Distrito Capital de Bogotá.
- Valor del contrato y valor de la multa: mayor a cincuenta mil (50.000) pesos para los contratos y mayor a cero (0) pesos para las multas.
- Fechas de inicio y fin del contrato: la fecha de inicio debe ser menor o igual al 01/08/2024 y la fecha de fin del contrato debe ser mayor o igual a la fecha de inicio del contrato.
- Tipo de adición a contrato: debe ser del tipo “Adición en el valor”.

Ahora, en cada uno de estos sets de datos se usará el modelo de Gradient Boosting para la predicción de tres variables objetivos, las cuales corresponden a tres indicadores de riesgo de corrupción que fueron seleccionados según las siguientes directrices que la OCDE describe en su caracterización de banderas rojas en procesos de contratación pública [7]:

Tabla 2.
Variables objetivo según Indicadores de la OCDE

Código OCDE del indicador	Descripción	Origen del indicador en los sets de Datos Abiertos
R069	Aprobación de órdenes de cambio innecesarias para aumentar el precio del contrato después de la adjudicación	¿El contrato tuvo adiciones en su valor, sí (True) o no (False)?
R046	El proveedor no figura en la lista de proveedores autorizados. Esto se traduce en que un proveedor haya sido sancionado	¿El contratista adjudicado para un contrato ha tenido sanciones por incumplimientos, sí (True) o no (False)?
R049	Alto número de adjudicaciones directas a un solo licitador	¿La entidad contratadora ha contratado más de una vez al mismo proveedor, sí (True) o no (False)?

4. MODELOS

Con el objetivo de predecir si el conjunto de las tres variables objetivo (contrato con adiciones en el valor, contrato adjudicado a proveedor sancionado y entidad con contratos a un mismo proveedor) se entrenó un modelo de Gradient Boosting Machine (GBM) con un enfoque de predicción multiclase.

El GBM pertenece a una categoría de modelos que combinan múltiples modelos "débiles" que individualmente no se ajustan bien a los datos, pero que juntos forman un modelo "fuerte" con alto rendimiento predictivo. Este modelo en particular construye de manera secuencial y codiciosa una serie potencialmente grande de árboles, donde en cada paso se ajusta un árbol de clasificación a los datos, pon-

derados por los residuos de los pasos previos. De esta forma, el GBM desarrolla un conjunto de árboles que logra un bajo error de generalización, al asignar mayor peso a las observaciones más difíciles de clasificar basándose en el desempeño de los árboles anteriores [8]. Debido a la naturaleza multiclase del problema se usó el modelo MultiOutputClassifier, el cual consiste en ajustar un clasificador por objetivo y se usa comúnmente para ampliar los clasificadores que no admiten de forma nativa la clasificación multiobjetivo [9].

Este modelo en específico fue escogido para la predicción del riesgo de corrupción debido al desbalance de los datos, detallado en la Tabla 3, y debido al desempeño que demostró tener el trabajo llevado a cabo por Gallego et al., quienes contaron con un desbalance de casos positivos similares al mostrado en el presente trabajo, y para quienes el modelo de Gradient Boosting Machine no presentó una diferencia de rendimiento significativa al usar técnicas de balanceo de datos [3]. También, la revisión hecha por Modrušan et al. señala el uso de este modelo y otros similares de aprendizaje supervisado para la identificación de indicadores y banderas rojas [6].

Tabla 3.
Balance de los casos positivos y negativos.

	Porcentaje de casos positivos	Porcentaje de casos negativos
Indicador de adiciones a contratos	4.67%	95.33%
Indicador de proveedor sancionado	0.43%	99.57%
Indicador de proveedor previamente contratado	79.47%	20.53%

Este modelo multiclase fue entrenado con los parámetros por defecto propios de cada modelo de Machine Learning de Scikit y con una división de los datos para entrenamiento de un 30 %.

5. CRITERIOS DE RENDIMIENTO

Para evaluar el rendimiento de un modelo de clasificación como lo es Gradient Boosting Machine, se usan las curvas ROC y AUC, adicional a esto, dado el desbalance de los datos, también se evaluará la curva de precisión-recuperación.

La curva ROC es una herramienta gráfica que muestra el desempeño de un modelo a lo largo de todos los posibles umbrales de clasificación. Su nombre completo, "Característica Operativa del Receptor", proviene de la terminología usada en la detección de radares durante la Segunda Guerra Mundial. La curva ROC se genera calculando la tasa de verdaderos positivos (TPR) y la tasa de falsos positivos (FPR) para cada umbral posible (o en puntos seleccionados), y luego graficando la TPR contra la FPR. Un modelo ideal tendría una TPR de 1.0 y una FPR de 0.0 [10].

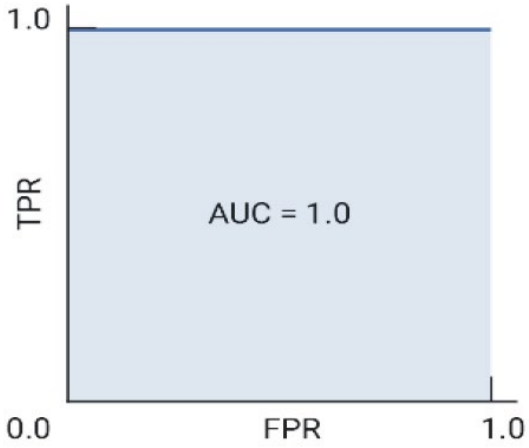

Figura 1.
ROC y AUC de un modelo hipotético perfecto [10].

ISSN 2322-8415 / Año 2024 / No. 24 - Vol. 14 / Páginas 61-70 Universidad Libre - Facultad de Ingeniería

El área bajo la curva ROC (AUC) representa la probabilidad de que el modelo, si se da un ejemplo positivo y negativo elegido al azar, clasificará el positivo mayor que el negativo. El modelo perfecto anterior, que contiene un cuadrado con lados de longitud 1, tiene un área bajo la curva (AUC) de 1.0. Esto significa que hay un 100% de probabilidades de que el modelo clasificará de manera correcta un ejemplo positivo elegido al azar más alto que un ejemplo negativo elegido al azar [10].

El AUC y la ROC funcionan bien para comparar modelos cuando el conjunto de datos es más o menos equilibrados entre clases. Cuando el conjunto de datos está desequilibrado, precisión-recuperación de las curvas de visión general (PRC) y el área debajo de ellas pueden ofrecer una mejor comparación visualización del rendimiento del modelo. Las curvas de precisión-recuperación se crean graficando la precisión en el eje Y y la recuperación en el eje X en todos umbrales [10].

6. RESULTADOS Y DISCUSIONES

Con los datos recolectados bajo las condiciones mencionados se realizó una limpieza básica eliminando filas vacías en los tres sets de datos e identificando las columnas con datos inconsistentes.

Las columnas del número de documento del proveedor y duración del contrato presentaron inconsistencias en sus datos: la primera no contenía sólo números de documento sino también cadenas de texto vacías, con correos electrónicos o combinados de números y caracteres especiales, por otro lado, la columna de duración del contrato presentaba valores nulos y con diferentes unidades de duración.

Esto se solucionó realizando una conversión del número de documento del proveedor a un dato de tipo numérico, aquellos registros que no pudieran ser convertidos fueron descartados del set de datos; en el caso de la duración del contrato se realizó una conversión a un tipo de dato numérico y se adoptó como unidad de medida los días, al igual que la primera columna, se descartaron los registros que no pudieran ser convertidos y aquellos que fueran mayores a 18250 días (50 años).

Con los datos consolidados en un set con los tres indicadores de riesgos de corrupción se realizó la división entre variables de entrada y variables objetivo, sobre las variables de entrada se realizó una codificación de las variables categóricas con LabelEncoder.

El conjunto de variables se dividió entre datos de entrenamiento y de testeo, los cuales se distribuyen de la siguiente manera para las variables objetivo.

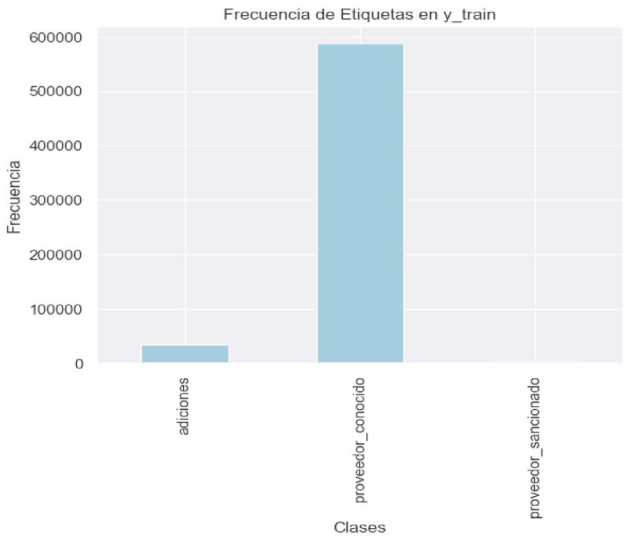

Figura 2.
Distribución de las variables objetivo para los datos de entrenamiento.

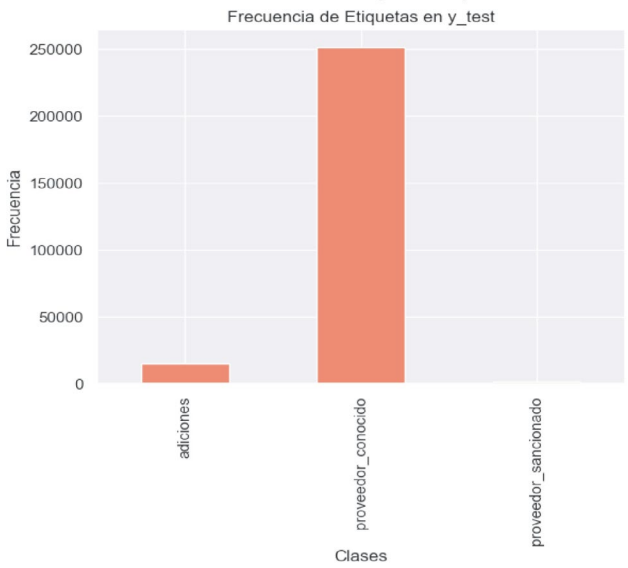

Figura 3.
Distribución de las variables objetivo para los datos de testeo.

Finalmente, las variables de entrenamiento fueron usadas como entrada para el modelo Gradient Boosting Machine, el cual realizó la predicción sobre los datos de testo resultando en los siguientes gráficos de curvas ROC - AUC y de precisión-recuperación.

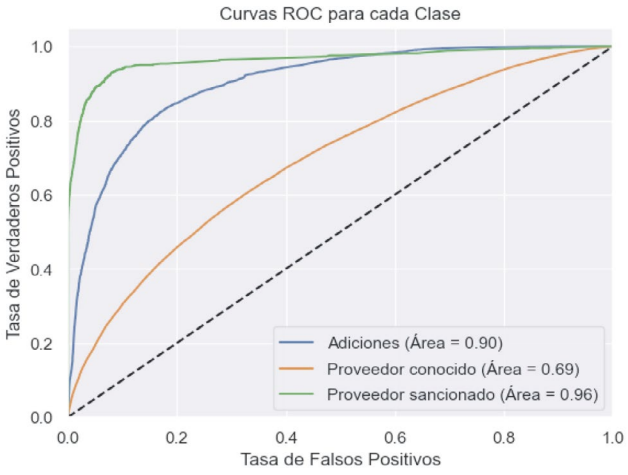

Figura 4.
Gráfica de curvas ROC para cada variable objetivo.

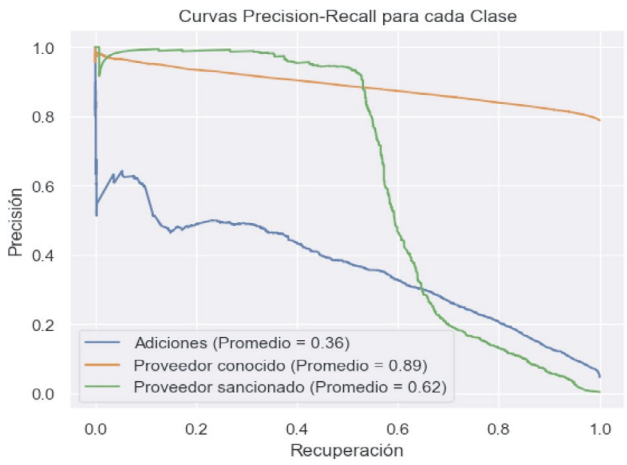

Figura 5
Gráfica de curvas de precisión-recuperación para cada variable objetivo.

Tabla 4.
Resumen de los parámetros de calificación.

Variable objetivo	Área bajo la curva ROC-AUC	Promedio del área bajo la curva precisión-recuperación
Adiciones	0.90	0.36
Proveedor conocido	0.69	0.89
Proveedor sancionado	0.96	0.62

Si bien las tres variables objetivo tuvieron un buen desempeño, estos resultados contrastan con los del promedio del área bajo la curva de precisión-recuperación, métrica que es más significativa en el presente estudio debido al desbalanceo de las variables objetivo dentro del set de datos.

7. CONCLUSIONES

Con el objetivo de evaluar el rendimiento de un modelo de Gradient Boosting Machine al momento de predecir el riesgo de corrupción en procesos de contratación pública se escogieron tres indicadores a predecir, los cuales fueron extraídos de los

ISSN 2322-8415 / Año 2024 / No. 24 - Vol. 14 / Páginas 61-70 Universidad Libre - Facultad de Ingeniería

tres sets de datos publicados por la plataforma Colombia Compra Eficiente en Datos Abiertos.

Un análisis realizado a estos sets de datos dio cuenta de las inconsistencias en la calidad de los datos, aun cuando estos fueron filtrados previamente según las conclusiones a las que llegaron los estudios de las auditorías citadas en este trabajo. Se realizó una estandarización y limpieza a dos de los campos con inconsistencias.

Los tres indicadores escogidos teniendo en cuenta la información disponible en Datos Abiertos y los estándares propuestos por la OCDE presentaron un desbalance significativo en valores positivos y negativos, no se realizó un proceso de balance manual de los datos debido a que dicho proceso no representó diferencias en el rendimiento en el modelo entrenado por Gallego et al. [3], el cual contó con un desbalance similar.

Un set de datos consolidado con los tres indicadores de riesgo fue usado en el entrenamiento del modelo bajo un enfoque de predicción multiclase, dando como resultado un modelo con áreas bajo la curva ROC superiores al 65% (ver Tabla 4), lo cual denota un buen rendimiento al momento de clasificar los valores positivos y negativos de cada indicador o variable objetivo.

Por otro lado, los resultados dados por el promedio bajo la curva de precisión-recuperación (ver Tabla 4) dan a conocer una mejor perspectiva del rendimiento del modelo teniendo en cuenta el desbalance de los datos de entrenamiento: puede observarse que las métricas para cada variable objetivo varían entre sí, siendo el indicador de Adiciones el que tuvo un rendimiento más bajo, el indicador de Proveedor Conocido teniendo uno regular y el indicador de Proveedor Sancionado teniendo el más alto.

Lo anterior sitúa al modelo Gradient Boosting Machine como una herramienta útil al momento de predecir indicadores de riesgo de contrato adjudicado a un proveedor previamente contratado por la entidad (Proveedor Conocido) o sobre un indicador de contrato adjudicado a un proveedor previamente sancionado por incumplimiento (Proveedor sancionado), sin embargo, el modelo no es útil al momento de predecir un indicador relacionado a adiciones realizadas a un contrato (Adiciones).

Por lo tanto, una mejor calidad en los datos de contratación pública en Datos Abiertos y el uso de otras fuentes para los indicadores podrían mejorar la precisión del modelo al incluir datos más limpios y con un menor porcentaje de desbalance, el cual podría ser útil en la implementación de sistemas de alerta temprana y en procesos de auditoría.

REFERENCIAS BIBLIOGRÁFICAS

[1] *E. A. Contreras Ussa,* "Auditoria de datos al sistema electrónico de contratación pública 'Colombia compra eficiente' sustentado en un modelo de aprendizaje no supervisado," instname:Universidad de Bogotá Jorge Tadeo Lozano, 2022, doi: 10/27843.

[2] *Corporación Transparencia por Colombia,* "Radiografía de los hechos de corrupción en Colombia," 2021, Accessed: Apr. 18, 2024. [Online]. Available: https://www.monitorciudadano.co/documentos/hc-informes/2021/Radiografia-2016-2021.pdf

[3] *J. Gallego, G. Rivero, and J. Martínez,* "Preventing rather than punishing: An early warning model of malfeasance in public procurement," Int J Forecast, vol. 37, no. 1, pp. 360–377, 2021, doi: 10.1016/j.ijforecast.2020.06.006.

[4] *J. Galvis, J. P. Marín, J. P. Garnica, H. Fonseca, H. Inga, and C. Cetina,* Guía para la identificación de riesgos de corrupción en contratación pública utilizando la ciencia de datos. 2021. [Online]. Available: https://ricg.org/wp-content/uploads/2021/12/Guia-para-la-identificacion-de-riesgos-de-corrupcion-en-CP-utilizando-la-ciencia-de-datos.pdf

[5] *OEA,* "Diagnóstico subregional sistema de compra y contratación pública," OEA - Organización de los Estados Americanos, 2021, [Online]. Available: https://www.oas.org/es/centro_noticias/comunicado_prensa.asp?sCodigo=C-053/21

[6] *N. Modrušan, K. Rabuzin, and L. Mršic,* "Review of Public Procurement Fraud Detection Techniques Powered by Emerging Technologies," International Journal of Advanced Computer Science and Applications, vol. 12, no. 2, 2021, doi: 10.14569/IJACSA.2021.0120272.

[7] *Open Contracting Partnership,* "Red flags for integrity: Giving the green light to open data solutions," 2016. Accessed: Aug. 27, 2024. [Online]. Available: https://www.open-contracting.org/resources/red-flags-integrity-giving-green-light-open-data-solutions/

[8] *J. H. Friedman,* "Greedy function approximation: A gradient boosting machine," The Annals of Statistics, vol. 29, no. 5, Oct. 2001, doi: 10.1214/aos/1013203451.

[9] *Scikit Learn,* "MultiOutputClassifier." Accessed: Aug. 27, 2024. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.multioutput.MultiOutputClassifier.html

[10] *Google,* "Clasificación: ROC y AUC." Accessed: Aug. 27, 2024. [Online]. Available: https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc?hl=es-419#:~:text=Receiver-operating%20characteristic%20curve%20(ROC),-The%20ROC%20curve&text=The%20ROC%20curve%20is%20drawn,then%20graphing%20TPR%20over%20FPR.