

# EL *MACHINE LEARNING* EN EL MODELO DE DETECCIÓN DE CÁRTELES EN LICITACIÓN PÚBLICA DE BAJARI Y SUMMERS

Yanela Ramírez, Alexis Ramos, Angela Jieza y Julio Aguirre

# El *machine learning* en el modelo de detección de cárteles en licitación pública de Bajari y Summers

Julio 2023



**Yanela Ramírez**

Estudiante de Economía en la Universidad del Pacífico.



**Alexis Ramos**

Estudiante de Economía en la Universidad del Pacífico.



**Angela Jieza**

Estudiante de último año de Economía en la Universidad del Pacífico.



**Julio Aguirre**

Ph.D. Economía, Universidad de San Andrés (Argentina). Profesor asociado del Departamento Académico de Economía de la Universidad del Pacífico (Lima, Perú). Sus áreas de especialización son la organización industrial, las políticas de competencia y regulación económica, y el desarrollo económico.

**Abstract:** Las prácticas de colusión en compras públicas o licitaciones estatales son un problema frecuente en países en desarrollo como el Perú, y su detección no deja de ser un reto. En las siguientes líneas, se presentan alcances de cómo el machine learning puede complementar al uso de modelos econométricos de detección (como el propuesto por Bajari y Summers, 2002).

\*\*\*

## I. INTRODUCCIÓN

El *bid rigging* es una práctica de cooperación realizada entre empresas para el suministro de bienes o servicios a través de la participación en procesos de licitación. Esto con el objetivo de presentar al ente público una oferta tentativa que resulte beneficiosa para los participantes de la colusión (Chang Tokushima & Drago Alfaro, 2022). De esta manera, la conducta del *bid rigging* puede adoptar diversas modalidades que se presentan a continuación:

### 1. Acuerdos de precio

Las empresas, en lugar de presentar ofertas competitivas, llegan a un acuerdo para elevar los precios y de esta manera fragmentar los contratos para distribuir de forma equitativa los beneficios. Un ejemplo es el que se presentó con las empresas Rheem Perú y Envases Metálicos, quienes acordaron incrementar





$CAP_{i,t}$  es una variable que mide la capacidad utilizada de la empresa  $i$  en el momento de licitar el proyecto  $t$ . Y  $MAXP_{i,t}$  es el porcentaje máximo de capacidad libre de todas las empresas en el proyecto  $t$ , excluida la empresa  $i$ .

$MDIST_{i,t}$  representa el mínimo de distancias para todas las empresas en el proyecto  $t$ , excluyendo la empresa  $i$ . Mientras que  $LMDIST_{i,t}$  es el  $\log(MDIST_{i,t} + 1.0)$ .

$CON_{i,t}$  mide la proporción de trabajo realizado (por volumen en dólares) por la empresa  $i$  en el estado donde se ubica el proyecto  $t$  antes de la subasta.

Y, por último, el término de perturbación  $\varepsilon_{j,t}$  es una variable que captura la variación de las ofertas no explicadas por los factores incluidos en la regresión. Parte de esta variación no explicada en las ofertas se debe a la información de costos privados que no es observable ni por el econométrico ni por otras empresas.

### Interpretación de las variables

La intuición económica básica nos permite una interpretación preliminar entre las variables. En primera instancia, se espera que tanto la oferta de la empresa  $i$  sea una función creciente respecto a la distancia de la empresa  $i$  al lugar del proyecto y al uso de la capacidad de la empresa. Dicho de otra manera, a medida que aumenta la distancia de la empresa  $i$  ( $DIST_{i,t}$ ), también lo debería hacer su costo.

Respecto a la variable  $CAP_{i,t}$  a medida que aumentan las ofertas de la empresa  $i$ , si todo lo demás se mantiene constante, entonces el valor de la capacidad libre aumentará porque una vez que la empresa  $i$  se vea completamente limitada en su capacidad, ya no tendrá la oportunidad de licitar en otros proyectos. Ahora, se esperaría que la variable  $CON_{i,t}$  guarde una relación negativa con las ofertas que presenta al Estado, lo cual es consistente con la hipótesis de que una mayor concentración de trabajo en un estado reduce los costos. Finalmente, las variables que interactúan entre empresas ( $MAXP_{i,t}$  y  $MDIST_{i,t}$ ) demostraría que, ante el aumento de la distancia o de la utilización de la capacidad de la empresa  $i$  o  $j$ , la competencia se suavice y por ende la empresa  $i$  aumente su oferta (Bajari y Summers, 2002).

### Prueba de Independencia Condicional

Para probar la independencia condicional entre ofertas se requiere que los residuos de las funciones de oferta para las empresas  $i$  y  $j$ ,  $\varepsilon_{i,t}$  y  $\varepsilon_{j,t}$  sea cero. Ello significa que los aspectos inobservables que afectan a las ofertas de una empresa  $i$ , no deberían guardar relación con los otros aspectos de una empresa  $j$ . Si el efecto resultante no es cero indicaría que alguna variable inobservable de la empresa  $i$  estaría afectando a la oferta de la empresa  $j$ .

Ahora, los residuos son una medida de la diferencia entre las ofertas reales y las ofertas que se podrían predecir utilizando la información pública disponible. Si las ofertas son independientes entre sí, no debería haber un patrón o relación constante en la diferencia entre las ofertas reales y previstas de diferentes empresas. Esta prueba se realiza para cada par de las once principales empresas. Como se explicó, la intercambiabilidad significa que cualquier empresa competitiva que se enfrente a un conjunto dado de condiciones de costos para sí misma y para sus rivales debe ofertar de la misma manera.

## **2. Intercambiabilidad**

Este criterio toma como base la eficiencia de los costos individuales de las empresas postoras. Así, si la estructura de costos y la oferta económica de las firmas restantes no se ve alterada, entonces se esperaría

que la conducta de la firma individual no se vea alterada pese a que los factores observados públicamente permuten. En otras palabras, bajo competencia, las empresas deberían presentar un comportamiento semejante bajo la misma estructura de costos entre ellas y en relación a sus rivales.

Los autores presentan un ejemplo de tres empresas: A, B y C, cuyos únicos costos en la industria son los costos de transporte, y el costo de una empresa es \$100,000 veces su distancia en millas<sup>5</sup> al sitio del proyecto.

**Tabla 1**  
**Costo para las firmas A,B y C**

Firmas	Distancia	Costo
A	10 millas	\$1.0 millones
B	12 millas	\$1.2 millones
C	13 millas	\$1.3 millones

Elaboración propia

Fuente: Bajari y Summers (2002)

¿Cómo deberían ofertar las empresas?

En primer lugar, como se observa en la tabla, A es la empresa que tiene menor costo, por lo que es la que tiene mayor probabilidad de ganar el contrato. Entonces, si las empresas no se coluden, el resultado de equilibrio en este caso sería que A oferte algo menos de \$1,2 millones y gane el contrato. Esta solución le permitiría a A obtener un beneficio extra y la potestad de ejecución del proyecto. Así mismo, en equilibrio, la empresa B debería presentar una oferta de \$1,2 millones, mientras que la empresa C debería ofertar \$1,3 millones. Tanto B como C evitarán ofertar menos para no arriesgarse a ganar el contrato y sufrir pérdidas financieras. Del mismo modo, tampoco ofrecerán más para intentar maximizar una posibilidad poco probable de ganar.

¿Cómo cambian las ofertas de equilibrio si A y B están en un cartel?

Asumimos que A y B pueden decidir formar un cartel de tal modo que el contrato salga favorable para A, pues esta tiene menores costos y conviene más, pues el cartel obtendrá mayores beneficios. En equilibrio, considerando que la empresa A no enfrenta el riesgo de una oferta más baja por parte de la empresa B, la empresa A debería hacer una oferta ligeramente inferior a \$1,3 millones. Mientras que la empresa C debería hacer una oferta de \$1,3 millones y la empresa B presentaría una oferta ficticia por encima de \$1,3 millones.

Pero, ¿qué sucede si las empresas A, B y C ofertan un segundo proyecto con distancias y costos como en la siguiente tabla?:

<sup>5</sup> El único costo que presenta cada empresa es el de transporte que equivale a 100 000 USD por cada milla de distancia que presenta la empresa respecto a la locación del proyecto.



**Tabla 2**  
**Costo para las firmas A,B y C en un segundo proyecto**

Firmas	Distancia	Costo
A	10 millas	\$1.0 millones
B	13 millas	\$1.3 millones
C	12 millas	\$1.2 millones

Elaboración propia

Fuente: Bajari y Summers (2002)

Si las empresas A y B llegan a un acuerdo, nuevamente deberían otorgar el proyecto a la empresa A debido a que A tiene costos más bajos que B (y, de hecho, la empresa B ya no puede hacer una oferta menor que C). Sin embargo, con el intercambio de costos entre B y C, la empresa A no hará la misma oferta que antes. En este caso, la empresa A debería hacer una oferta ligeramente inferior a \$1,2 millones, ya que C, que no está involucrada en el acuerdo, hará una oferta de \$1,2 millones.

**Tabla 3**  
**Falta de Intercambiabilidad con licitación colusoria**

Firma	Costo del 1° proyecto	Oferta Competitiva para el 1° proyecto	Oferta Colusiva para el 1° proyecto	Costo para el 2° proyecto	Oferta Competitiva para el 2° proyecto	Oferta Colusiva para el 2° proyecto
A	\$1.0 millones	\$1.19 millones	\$1.29 millones	\$1.0 millones	\$1.19 millones	\$1.0 millones
B	\$1.2 millones	\$1.2 millones	Oferta falsa > \$1.29 millones	\$1.3 millones	\$1.3 millones	\$1.3 millones
C	\$1.3 millones	\$1.3 millones	\$1.3 millones	\$1.2 millones	\$1.2 millones	\$1.2 millones

Elaboración propia

Fuente: Bajari y Summers (2002)

En esta tabla, se puede observar que las ofertas competitivas son intercambiables. Cuando dos empresas competidoras tienen el mismo costo y enfrentan los mismos costos por parte de sus rivales realizan la misma oferta. En el caso de la empresa A, sus ofertas competitivas son idénticas para el primer y segundo proyecto debido a que el costo de su rival más cercano (empresa B en el primer proyecto y empresa C en el segundo proyecto) es el mismo. Por lo tanto, A realiza la misma oferta en ambos proyectos. Por otro lado, las ofertas competitivas de B y C se intercambian a medida que se intercambian sus costos entre los dos proyectos. En el primer proyecto, B tiene un costo de \$1,2 millones y los costos de sus rivales son \$1,0 millones y \$1,3 millones, por lo que B ofrece \$1,2 millones. Al pasar al segundo proyecto, las posiciones de B y C se "intercambian" y la empresa C realiza una oferta de \$1,2 millones.

Sin embargo, en la Tabla 3 también se puede observar que la intercambiabilidad de costos y ofertas se rompe cuando la oferta es colusoria. En este caso, la empresa A realiza una oferta de \$1,29 millones en el primer proyecto y \$1,19 millones en el segundo proyecto, a pesar de que el costo de su "rival" más cercano permanece sin cambios. Lo que ha cambiado es la participación del rival con el costo más cercano en el cartel. Además, con la oferta colusoria, las ofertas de las empresas 2 y 3 ya no se intercambian cuando se intercambian sus costos.

### 3. Carrera de caballos

Para fortalecer la validez de las pruebas iniciales, Bajari y Ye proponen agregar una tercera prueba. Las primeras pruebas pueden detectar posibles cárteles al identificar empresas que presentaron ofertas que no son independientes o intercambiables en ciertas condiciones. La tercera prueba consiste en comparar un modelo competitivo con modelos de colusión que involucran a los cárteles potenciales identificados en las pruebas anteriores. Esta prueba adicional, conocida como “carrera de caballos”, busca confirmar la existencia de colusión y distinguirla de un comportamiento competitivo. El proceso consiste en el análisis de la distribución de los márgenes de ganancia en las empresas en diversos escenarios a través de datos observables. Su objetivo es comparar la probabilidad de que un proceso colusorio produzca una distribución de ganancia significativa, mientras que los modelos competitivos tenderían a predecir un comportamiento opuesto.

Esta prueba utiliza información sobre la estructura de costos y márgenes típicos observados en la industria. Para obtener esta información, se recurre a diversas fuentes, como empresas que no están bajo sospecha de colusión, consultores de ingeniería que realizan estimaciones de costos para proyectos similares y registros internos de licitadores obtenidos por las autoridades. Con esta información en mano, se lleva a cabo un análisis empírico mediante una “carrera de caballos” entre diferentes modelos competitivos y modelos de colusión. El objetivo es determinar la probabilidad de que cada modelo explique el comportamiento de las empresas oferentes, considerando las ofertas observadas y la información previa sobre costos y márgenes de la industria.

Estas pruebas de colusión fueron aplicadas por Bajari y Ye a un conjunto de datos de licitación de empresas constructoras en Minnesota, Dakota del Norte y Dakota del Sur durante los años 1994-1998, específicamente para un tipo de reparación de carreteras conocido como revestimiento de sellado. El trabajo de los autores amplía y es coherente con el análisis realizado en otros documentos empíricos recientes sobre manipulación de licitaciones, como los trabajos de Porter y Zona, así como el de Pesendorfer.

Aunque estas pruebas pueden ser útiles para detectar comportamientos sospechosos en las licitaciones, es importante tener en cuenta que ningún método para detectar colusión es infalible y que estas pruebas tienen sus limitaciones. Además, aunque pueden señalar desviaciones del comportamiento competitivo, no determinan directamente si una desviación es resultado de un acuerdo ilegal, sino que más bien indican una posible forma de colusión tácita en la que las empresas actúan de forma independiente pero conscientes de los efectos de su comportamiento en relación con las demás. A pesar de estas consideraciones, estas pruebas sí pueden servir como un primer paso para identificar licitaciones sospechosas, y así justificar una investigación y análisis más exhaustivos.

## III. LIMITACIONES DEL MODELO BAJARI Y SUMMERS

Si bien el modelo de Bajari y Summers (2002) es aplicado en diversos análisis de prácticas colusorias, se ponen en evidencia algunas limitaciones al emplear sus criterios de análisis. Por un lado, una firma puede actuar de manera anticipada y presentar ofertas que cumplan con la independencia condicional y la intercambiabilidad. De esta forma, basándose únicamente en esta evidencia, no es posible descartar la posibilidad de que las ofertas en cuestión provengan de un proceso de licitación pública competitivo (Huber e Imhof, 2019). En este sentido, el modelo de Bajari y Summers puede tender a presentar un alto número de falsos negativos, puesto que las pruebas asumen la presencia de competencia cuando no existe (Imhof y Wallimann, 2021).

Además, si las funciones de oferta o de costos no están bien establecidas debido a falta de información, existe una mayor posibilidad de error en los resultados de las pruebas (García Rodríguez et al., 2022).



Por último, existen factores exógenos como cambios en las condiciones del mercado o regulaciones gubernamentales que pueden afectar la estructura de las ofertas; de modo que los resultados no se vean reflejados en la independencia condicional.

## IV. EL MODELO DE MACHINE LEARNING

Ante las limitaciones mencionadas, es posible complementar las ideas del modelo de Bajari y Summers (2002) con el apoyo del machine learning. En principio, el *machine learning* (o el aprendizaje automático) se define como el campo de estudio que permite a los ordenadores la capacidad de aprender sin necesidad de algún tipo de programación explícita. El modelo opera a través de un proceso de entrenamiento que consiste en enseñarle a reconocer ciertas características y construir un modelo de aprendizaje a través de los datos que ingresemos por medio de dos etapas principales (Nasteski, 2017).

### 1. *Data Training* (Datos de entrenamiento)

En esta primera etapa, el *machine learning* se expone a un conjunto de datos de entrenamiento que son usados para entrenar a un algoritmo o modelo de aprendizaje automático con la finalidad de predecir un resultado concreto. Los datos consisten en ejemplos de entrada con sus respectivas salidas, y etiquetas (labels) que deben ser supervisadas por las personas para elegir las características de los datos que se desea que la máquina reconozca. En cambio, un aprendizaje no supervisado implica el uso de datos no etiquetados para encontrar ciertos patrones como agrupaciones de datos o inferencias.

Durante este proceso, el modelo ajusta sus parámetros para aprender a mapear de forma correcta ciertas características de entrada de acuerdo con salidas esperadas. Todo ello a través de algoritmos de optimización que minimizan los errores. Por ejemplo, uno puede decidir entrenar al modelo para que reconozca las imágenes de computadoras y lapiceros. Para ello, los datos de entrenamiento contendrían imágenes de computadoras con la etiqueta de “COMPUTADORA” y los lapiceros con la etiqueta de “LAPICERO”.

### 2. *Reinforcement Learning* (Refuerzo)

El refuerzo es una forma de entrenamiento basada en la interacción de un agente con un entorno para planear estrategias efectivas en base a la experimentación. En esta etapa, el modelo emprenderá las acciones que repetirá e intentará reforzar de acuerdo con las recompensas que obtenga, tanto positivas como negativas. En otras palabras, el *reinforcement learning* consiste en un sistema de construcción autónomo que permita a la inteligencia seguir el camino correcto y evitar aquellos en donde se presentan los errores.

En general, la utilidad del *machine learning* es que permite enseñar a las máquinas a manejar bases de datos para el estudio de algoritmos y modelos estadísticos de forma que, una vez que un algoritmo logre aprender sobre el uso de los datos, este pueda hacer su trabajo de forma automática (Mahesh, 2020).

El algoritmo y la inteligencia artificial a través de las redes neuronales y el aprendizaje profundo (*deep learning*) permiten detectar casos de anomalía<sup>6</sup> que difícilmente podrían ser detectadas por una entidad pública. Esto debido a su facilidad para analizar datos históricos e identificar ciertos patrones. En términos específicos, las

---

<sup>6</sup> Los casos de anomalía corresponden a escenarios o eventos que difieren de forma significativa de lo esperado o de la norma de un determinado contexto.

colusiones pueden verse facilitadas por estudios en los algoritmos de supervisión, algoritmos paralelos, de señalización o incluso aquellos de aprendizaje (OECD, 2022). De esta manera, se pueden destacar algunas ventajas que conlleva la implementación del machine learning en los procesos de detección de bid rigging:

## V. VENTAJAS DEL MACHINE LEARNING EN LOS PROCESOS DE LICITACIÓN

### 1. Ganancias en eficiencia

La OCDE señala que eliminar la manipulación de ofertas puede resultar en una reducción de hasta el 20% en los precios de las adquisiciones. Por lo tanto, es crucial desarrollar métodos proactivos para detectar conspiraciones de manipulación de ofertas, lo cual es de gran relevancia para las agencias de competencia y adquisiciones a nivel global (OCDE citado por Imhof y Wallimann, 2021). Dado que los procesos de licitación requieren el análisis y procesamiento de grandes cantidades de datos, el machine learning puede recopilar datos que son encontrados fácilmente en los registros oficiales y las ofertas públicas. De esta manera, habría una mayor eficiencia en términos de tiempo y recursos, además de que su aplicación discrecional permitiría que estos procesos pasen desapercibidos por los cárteles.

### 2. Aplicabilidad a diversos contextos

Debido a que el machine learning posee la capacidad de trabajar de forma automática una vez que logra aprender el algoritmo, es posible que estime diversos modelos aplicables según el contexto en el que se encuentre o los datos que posea (Mahesh, 2020). Además, puede analizar grandes volúmenes de datos relacionados con licitaciones, como precios, ofertas y características de los licitantes, para identificar patrones que indiquen posibles prácticas colusorias

### 3. Menores probabilidades de error

Uno de los principales inconvenientes mencionados del modelo Bajari y Summers es que su modelo econométrico tiende a producir un alto número de falsos positivos. En otras palabras, las pruebas identifican ausencia de competencia en el mercado pese a que la clasificación acertada es que el mercado está presentando presencia de *bid rigging*. Sin embargo, el machine learning permite que las predicciones tengan un rendimiento decente debido a que los parámetros del modelo pueden ser modificados de acuerdo con las necesidades de los tribunales (Huber et al., 2019).

Imhof y Wallimann (2021), por otro lado, han desarrollado un modelo que demuestra la factibilidad de detectar acuerdos colusivos en subastas de adquisiciones mediante un método novedoso basado en el original método de *screening*. Su enfoque basado en colusiones, permite identificar directamente a las empresas involucradas en los cárteles y puede detectar tanto cárteles completos como incompletos que manipulan las ofertas. Además, si el modelo se calibra correctamente, también puede identificar cárteles parciales, es decir, aquellos que se dedican a la manipulación de ofertas en áreas o tipos específicos de contratos. Esta investigación representa un avance significativo en la detección y prevención de prácticas colusorias en subastas de adquisiciones.

### 4. Procesos de licitación *ex-ante*

Los modelos predictivos a través del machine learning permiten examinar posibles casos de colusión en procesos de manipulación de licitaciones. Otra limitante del modelo de Bajari y Summers (2002) es la incapacidad para detectar casos de posible colusión cuando una firma actúa de manera anticipada y presenta

ofertas que cumplan con la independencia condicional y la intercambiabilidad. Por ello, en el caso de que el organismo encargado de la supervisión no cuente con datos suficientes o incluso ninguno, puede optar por usar parámetros genéricos del modelo de Huber e Imhof (2019) para detectar casos potencialmente sospechosos.

## VI. LA APLICACIÓN DE MACHINE LEARNING EN LOS PROCESOS DE LICITACIÓN

### El mercado de construcciones de Suiza

En el estudio previamente mencionado por Huber e Imhof (2019), se analiza el mercado de construcción y mantenimiento de carreteras de Suiza mediante una muestra de 584 licitaciones con casi 3.800 ofertas. El rendimiento del modelo con filtros estadísticos fuera de la muestra logró clasificar hasta un 84% los procesos de licitación entre colusorias y no colusorias.

Huber e Imhof (2019) nos explican que cuatro cárteles en Suiza, ubicados en Ticino y St. Gallen, manipularon las licitaciones del sector construcción y mantenimiento de carreteras. Durante un periodo de tiempo, conocido como periodo de cártel, estas empresas coludieron y manipularon licitaciones con el fin de ganar contratos a precios inflados. Después, en el periodo post cártel, se retornó a la competencia justa. Esto dejó un patrón comparable entre cantones que sería de utilidad para la realización de un método de machine learning.

En el caso suizo, se emplearon dos enfoques basados en aprendizaje automático para predecir carteles involucrados en manipulaciones de licitaciones. El primero es la técnica de regresión de Lasso, una variante de regresión logística, por otro lado, el segundo es un método que resultó del ensamble de árboles de regresión<sup>7</sup>, ensacado<sup>8</sup>, bosques aleatorios<sup>9</sup> y redes neuronales<sup>10</sup>.

El enfoque de regresión de Lasso consistió en partir el conjunto de datos en dos subconjuntos: una muestra de entrenamiento y una muestra de prueba. Posteriormente, se estimó la existencia de un cartel usando una función que considera predictores como variables originales. Finalmente, se estableció una relación funcional flexible entre las variables y la presencia del cartel.

En relación al enfoque de conjunto en el análisis de la existencia de cárteles, se utilizó una combinación ponderada de árboles de regresión con ensacado, bosques aleatorios y redes neuronales. Estos algoritmos se basan en la aplicación de técnicas basadas en árboles que dividen recursivamente los datos en submuestras con el objetivo de minimizar las diferencias cuadradas entre las incidencias reales de los cárteles y las probabilidades de colusión dentro de dichas submuestras.

---

7 Los árboles de regresión son árboles de decisión diseñados específicamente para predecir valores continuos de variables objetivo en lugar de etiquetas de clase discretas, utilizando criterios modificados de división y detención.

8 “Bagging” o ensacado es una técnica en el aprendizaje automático que mejora la precisión de los modelos al combinar predicciones de múltiples modelos perturbaciones en diferentes subconjuntos de datos, reducir la conveniencia en las predicciones y mejorar el rendimiento general.

9 “Random forest” es un conjunto de árboles de decisión que combinan sus predicciones para realizar predicciones precisas. Se beneficia de la baja correlación entre los árboles y depende de características significativas para obtener buenos resultados.

10 Las redes neuronales son un subconjunto del aprendizaje automático que imitan la comunicación entre las neuronas biológicas. Son fundamentales en los algoritmos de aprendizaje profundo y están inspiradas en el cerebro humano.

## El Cártel de Okinawa

El estudio se centra en el desmantelamiento del cartel del mercado de construcciones y contratos de ingeniería civil que manipulaba licitaciones en Okinawa. Los participantes del cartel coordinaron sus ofertas, eligieron un ganador y presentaron ofertas falsas, aprovechando la ventaja de que las condiciones geográficas dificultaban la entrada de nuevos competidores al mercado. El caso fue desmantelado por la Comisión de Comercio Justo de Japón (JFTC) en junio del 2005. El cártel involucrado había recurrido a prácticas colusorias y, para su detección, se usaron los patrones presentados en Huber e Imhof (2019) aplicado al país suizo.

Después de que la Comisión de Comercio Justo investigara e impusiera sanciones, se realizaron cambios en el sistema legal y de adquisiciones. Esto incluyó un aumento en las multas y la introducción de un programa de clemencia, así como la modificación del proceso de invitación y la no divulgación de información antes de la licitación. Este estudio demuestra que combinar las *screens* estadísticas y el *machine learning* ayuda a desmantelar cárteles dado que, desde un punto de vista cualitativo, las licitaciones suizas se comportaron de forma similar a las japonesas. Huber, Imhof, & Ishii (2020) utilizan métodos como el *Random Forest*<sup>11</sup> y *Super Learner*<sup>12</sup>.

El método de machine learning usado en la detección de carteles en Okinawa se basó en el uso de la regresión Lasso, el algoritmo *Random Forest* (máquina de vectores de soporte<sup>13</sup>) y el modelo *Super Learner*. Estos algoritmos se aplicaron de forma conjunta para mejorar la precisión de las predicciones.

La regresión Lasso consiste en utilizar una variante de regresión logística que excluye los coeficientes de baja capacidad predictiva, permitiendo obtener modelos más escasos con solo las variables más potentes en términos de predicción.

Por otro lado, el *Random Forest* es un algoritmo que combina múltiples árboles de decisión para poder predecir resultados. Estos árboles se entrenan con muestras aleatorias de datos. Los árboles se combinan para poder hallar una predicción final. De esta forma, se maximiza el provecho de la información de cada árbol. Asimismo, las máquinas de vectores de soporte buscan un hiperplano que separe los datos de entrenamiento, y usa la función Kernel para mapear datos, lo que mejora la capacidad para discernir entre la información.

Por último, el *Super Learner* combina los resultados de varios algoritmos, es decir, es la combinación de los métodos mencionados anteriormente para la predicción final.

## El mercado de Brasil, Italia y Estados Unidos

El estudio de García Rodríguez et al. (2022) se centró en probar la aplicación de once algoritmos distintos de machine learning a partir de seis conjuntos de datos de distintos mercados para diversos países que habían presentado un caso de *bid rigging*: Brasil, Italia, y Estados Unidos<sup>14</sup>:

---

11 El Random Forest es un enfoque de aprendizaje automático basado en el concepto de árboles de decisión para la toma de decisiones, para predecir una variable binaria conformada por distintas variables explicativas.

12 El Super Learner es un modelo estadístico no lineal usado en el aprendizaje automático para capturar patrones no lineales y relaciones difíciles de detectar en los datos. Por medio de un entrenamiento se ajustan sus conexiones y pesos para lograr predecir de manera precisa las salidas deseadas.

13 Support Vector Machine (SVM) es un algoritmo popular de aprendizaje supervisado utilizado para clasificación. Encuentra un límite de decisión óptimo llamado hiperplano para separar los puntos de datos en diferentes clases. SVM se basa en la identificación de vectores de soporte, puntos/vectores extremos que definen el hiperplano.

14 En el documento original se mencionan en total 5 países: Brasil, Italia, Estados Unidos, Japón y Suiza. Sin embargo, estos dos últimos países han sido abordados en los párrafos previos.

En Brasil, durante los años 2002 y 2013, la empresa petrolera brasileña Petrobras, empresa controlada por el Estado para su cotización en bolsa, fue objeto de diversas manipulaciones en licitaciones de proyectos en infraestructura. En general, las pruebas demostraron que se habían realizado pagos indebidos hacia empresas contratistas. La investigación federal, también llamada operación "Autolavado" (*Lava Jato*) demostró que 16 de las empresas petroleras más grandes de Brasil se habían puesto de acuerdo para coludirse. La operación, dirigida por la Policía Federal de Brasil, recibió dicho nombre dado que las empresas hacían uso de redes de lavandería y estaciones de servicio para mover grandes sumas de origen ilícito.

Por otro lado, los datos de Italia comprenden subastas de construcción de carreteras del municipio de Turín donde varias empresas, entre los años 2000 y 2003, fueron acusadas de colusión. Finalmente, para el año 2008, el Tribunal de Justicia de Turín castigó a un total de 95 empresas que operaban en diversos cárteles que habían adjudicado contratos con éxito. En general, menos del 10% de las empresas ganaron más del 80% de las subastas ofrecidas.

El caso de Estados Unidos destacó por involucrar contratos de adquisición en el mercado de lácteos para los colegios del estado de Ohio entre los años 1980 y 1990. Los funcionarios del distrito escolar solicitaron ofertas en contratos anuales de diversas empresas. Para la subasta se seleccionó al postor más bajo para que suministre su producto en los colegios durante el siguiente año. Sin embargo, en 1993 dos lecherías de Ohio confesaron haber manipulado las subastas en los 80. De esta manera, al recopilar los datos necesarios, se detectaron en total 30 lecherías que habían participado de la colusión que resultó en un aumento del 6.5% en los precios de sus productos.

De acuerdo con los datos recopilados por los autores, los resultados del estudio mostraron que, pese a los datos limitados, los algoritmos habían presentado buenos rendimientos. Incluso, los algoritmos bien entrenados y actualizados podrían ir adaptándose de forma automática a subastas más recientes. En general, los once algoritmos de *machine learning* se probaron en diversas métricas con resultados satisfactorios de detección de procesos colusivos que oscilaron entre un 81% y 95%, además de presentar una precisión equilibrada superior al 73%<sup>15</sup>.

## **Limitante del Software y conclusiones**

Si bien el machine learning resulta atractivo para su incorporación en los procesos de colusión debido a sus aplicaciones a la realidad, debemos considerar que los algoritmos, al comportarse de forma automatizada y de manera artificial, pueden incurrir en un análisis que implique colusión ilegal cuando se podría dar el caso de alguna práctica coordinada que termine siendo beneficiosa para el mercado. Por tanto, se sugiere que el machine learning sea implementado con rigurosidad y siempre de la mano de la teoría económica.

En este sentido, se debe tomar en consideración que el uso de los algoritmos y softwares de la inteligencia artificial sirven como un soporte para ahorrar tiempos y costos, pero al final su función no puede suplantar en su totalidad al análisis de un agente público.

El uso del machine learning en casos de bid rigging ofrece beneficios significativos al combinar la teoría económica con la capacidad de análisis de grandes volúmenes de datos. Al aplicar algoritmos y técnicas de machine learning, las autoridades encargadas pueden detectar patrones de comportamiento sospechosos e identificar posibles

---

15 Dentro del porcentaje se excluyen los resultados variables del caso de Estados Unidos debido a una menor efectividad en sus resultados por la menor cantidad de postores y ofertas encontradas.

cárteles o coaliciones en licitaciones. Esto fortalece la capacidad de las autoridades para prevenir y combatir la colusión en las licitaciones, lo que a su vez fomenta la competencia en el mercado y protege los intereses de las organizaciones y los consumidores. Al evitar la manipulación de ofertas y promover licitaciones justas y transparentes, se reducen los precios inflados y se maximiza el valor obtenido en las adquisiciones.

Sin embargo, es importante destacar que el machine learning no debe ser considerado como una solución única. Es necesario complementar su uso con una combinación de herramientas legales, regulatorias y de supervisión. Además, es crucial que las autoridades cuenten con personal capacitado y especializado en el análisis de datos y en la interpretación de los resultados obtenidos mediante machine learning.

En términos de recomendaciones, se sugiere que las autoridades inviertan en la capacitación de su personal y en el desarrollo de alianzas con expertos en machine learning y análisis de datos. Asimismo, es importante establecer políticas y regulaciones claras que promuevan la transparencia en los procesos de licitación y que disuadan la colusión. Además, se debe fomentar la colaboración y el intercambio de información entre las autoridades a nivel nacional e internacional para mejorar la detección y prevención de prácticas colusorias en licitaciones.



## REFERENCIAS

Bajari, P., & Summers, G. (2002). *Detecting collusion in procurement auctions*. *Antitrust LJ*, 70, 143.

Chang, J. & Drago, M. (2022). *Bid rigging o concertación en licitaciones: ¿Qué, cuándo y cómo se produce un hard core cartel bajo esta modalidad*. *THEMIS Revista de Derecho*, 81, 119–134. Recuperado de <https://doi.org/10.18800/themis.202201.009>

García Rodríguez, M. J., Rodríguez-Montequín, V., Ballesteros-Pérez, P., Love, P. E. D., & Signor, R. (2022). *Collusion detection in public procurement auctions with machine learning algorithms*. *Automation in Construction*, 133. <https://doi.org/10.1016/j.autcon.2021.104047>

Huber, M., Imhof, D. (2019). *Machine learning with screens for detecting bid-rigging cartels*. Recuperado de [https://www.sciencedirect.com/science/article/abs/pii/S0167718719300219?casa\\_token=hFFo\\_N9eAHoAAAAA:vHLolvc-VuY4wg5g1kTleSjqfYeDDc57B8SZTm1-81\\_bOWshNox5\\_DdW7RZucbE37Gq8mjPVW5Zw](https://www.sciencedirect.com/science/article/abs/pii/S0167718719300219?casa_token=hFFo_N9eAHoAAAAA:vHLolvc-VuY4wg5g1kTleSjqfYeDDc57B8SZTm1-81_bOWshNox5_DdW7RZucbE37Gq8mjPVW5Zw)

Huber, M., Imhof, D., & Ishii, R. (2020). *Transnational machine learning with screens for flagging bid-rigging cartels*. Fribourg: Université de Fribourg.

Imhof, D., Wallimann, H. (2021). *Detecting bid-rigging coalitions in different countries and auction formats*. Recuperado de [https://www.sciencedirect.com/science/article/pii/S0144818821000405?casa\\_token=-quwHFy8iMoAAAAA:rShlh9w10HPEW5HaGFj4Sl7J4mMUPLyqnHSQG1TBDA5OyM2F8IP4Iz0htDW3TxswXRcgjSr4bgCD](https://www.sciencedirect.com/science/article/pii/S0144818821000405?casa_token=-quwHFy8iMoAAAAA:rShlh9w10HPEW5HaGFj4Sl7J4mMUPLyqnHSQG1TBDA5OyM2F8IP4Iz0htDW3TxswXRcgjSr4bgCD)

Mahesh, B. (2020). *Machine Learning Algorithms-A Review*. *International Journal of Science and Research*, 9(1). <https://doi.org/10.21275/ART20203995>

Nasteski, V. (2017). *An overview of the supervised machine learning methods*. *Horizons*. b, 4, 51-62.

OECD (2022), *Data Screening Tools in Competition Investigations*, OECD Competition Policy Roundtable Background Note, [www.oecd.org/daf/competition/data-screening-tools-in-competitioninvestigations-2022.pdf](http://www.oecd.org/daf/competition/data-screening-tools-in-competitioninvestigations-2022.pdf).



Este documento se encuentra sujeto a los términos y condiciones de uso disponibles en nuestro sitio web:  
<http://www.centrocompetencia.com/terminos-y-condiciones/>

**Cómo citar este artículo:**

Yanela Ramírez, Alexis Ramos, Angela Jieza y Julio Aguirre, "El *machine learning* en el modelo de detección de cárteles en licitación pública de Bajari y Summers", *Investigaciones CeCo* (julio, 2023),  
<http://www.centrocompetencia.com/category/investigaciones>

Envíanos tus comentarios y sugerencias a [info@centrocompetencia.com](mailto:info@centrocompetencia.com)  
CentroCompetencia UAI – Av. Presidente Errázuriz 3485, Las Condes, Santiago de Chile