

# Limitaciones a la competencia en procesos de contratación pública de vías primarias

## 1. Introducción

El gasto público tiene como propósito poner a disposición de las personas los bienes, obras y servicios que están a cargo de las Entidades Estatales y actividades tales como el pago de deuda, salarios, pensiones, subsidios, y transferencias a nivel nacional. En Colombia el Sistema de Compra Pública es la estructura utilizada para tomar las decisiones sobre el gasto público y adicionalmente media las relaciones entre los actores del mercado, Entidades Estatales y autoridades competentes (Colombia Compra Eficiente, 2016). Entre el 2014 y el 2018 los contratos de obra pública en el país superan el valor de 97 billones de pesos y representan el 23% del total de recursos contratados, siendo superado únicamente por los contratos de prestación de servicios.

**TIPO DE CONTRATO SEGÚN PORCENTAJE DEL VALOR TOTAL CONTRATADO**

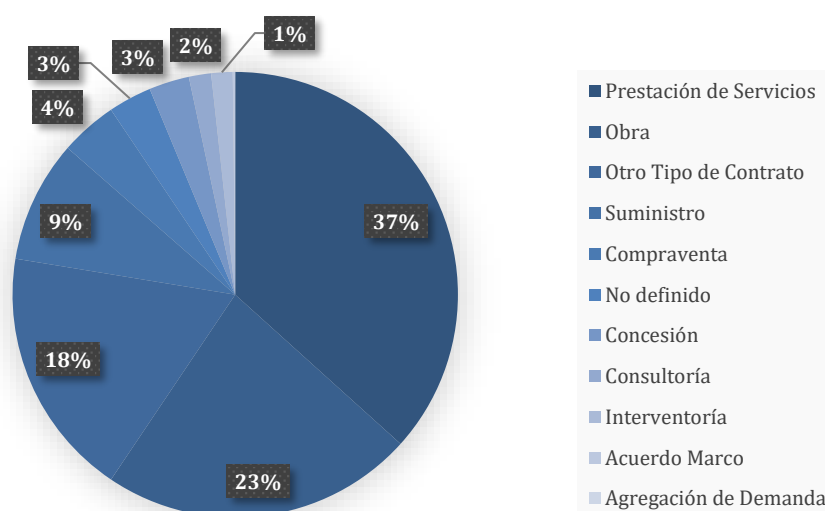


Gráfico 1. Fuente: Elaboración propia con base en SECOP I

Dentro de los contratos de obra pública resaltan los de vías primarias, las cuales conectan las principales áreas de producción y consumo del país, dado que estos son procesos de grandes cuantías, largos plazos de ejecución y su implementación incide en el desarrollo económico

del país. La ejecución eficiente de los recursos en estos proyectos tiene una gran importancia para la competitividad del país, pues permite la conexión entre regiones y facilita las dinámicas económicas.

Sin embargo, a pesar de que en el país existe un amplio marco normativo que busca proteger la transparencia y equidad en los procesos de compra pública, los numerosos escándalos de corrupción evidencian que no siempre ocurre de esta manera. Según las estimaciones de Transparencia por Colombia (2019), entre enero de 2016 y julio de 2018 en los 327 casos de corrupción que fueron reportados, se puso en riesgo alrededor de 17,9 billones de pesos, valor equivalente a la mitad del presupuesto destinado a educación en 2018. Dentro de las múltiples prácticas en ocasiones asociadas a corrupción se encuentran las que afectan la competencia del mercado y provocan que los procesos tengan sobrecostos en su ejecución los cuales, según las estimaciones de la OCDE, se encuentran en el orden del 10 al 20%. Frecuentemente, los contratos relacionados con la construcción de vías son “objeto de señalamientos en los medios de comunicación y en los gremios de la construcción por restricción de la competencia” (Zuleta y Caro, 2020); esto, sumado al impacto que tiene su implementación, hace que se necesiten revisar con más detalle. En consecuencia, en este trabajo se analizaron contratos de vías primarias con el propósito de identificar patrones que puedan indicar posibles falencias en la competitividad que se espera tengan estos procesos. Para esto, se construyó una base de datos que recopila la información resultante de la identificación y caracterización de los proponentes que participaron en una muestra de estos procesos. Posteriormente, mediante la evaluación de métricas se seleccionó un algoritmo de Machine Learning para identificar casos que podrían indicar falta de competencia en contratos de vías primarias. Por último, se desarrolló un sistema de alertas que permite establecer la probabilidad de que, dadas sus características, un proponente pueda estar relacionado en prácticas que afecten la competitividad de los procesos de contratación.

## **2. Justificación**

Según el barómetro global de corrupción para América Latina y el Caribe 2019, en Colombia el 94% de las personas consideran que uno de los problemas más grandes del país es la corrupción (Transparencia por Colombia). Adicionalmente, la corrupción afecta el desarrollo del país y el acceso a los bienes y servicios que se encuentran a cargo del Estado, puesto que

los recursos públicos que deberían ser invertidos en áreas como educación, salud o infraestructura terminan en manos de terceros.

De acuerdo con el informe de la plataforma Monitor Ciudadano de la Corrupción, entre enero de 2016 y julio de 2018 se presentaron 327 casos de corrupción donde el 15% pertenecían al sector de infraestructura (2019). Algunos de estos casos tuvieron una amplia cobertura mediática como el caso Odebrecht en la Ruta del Sol II-III, donde se encontraron involucrados políticos y grandes empresas. Este caso tuvo un gran impacto pues no solo afectó la construcción de una obra pública, sino que permeó un gran número de procesos y entidades tanto públicas como privadas. Todo esto lleva a que la corrupción tenga un impacto negativo en la confianza en el Gobierno por parte de los ciudadanos y pone en duda el trabajo desarrollado por las entidades estatales, lo cual se refleja en cifras como las que se reportan en el Barómetro Colombiano de la Reconciliación, iniciativa de la Usaid y Acdi/Voca, donde el 88.9% de los colombianos encuestados no confían en el Gobierno Nacional (2019).

La corrupción se presenta en muchas formas, entre las que se encuentran las prácticas que restringen la competencia donde algunos participantes obtienen ventajas de forma deshonestas. Una de estas prácticas es la colusión, en la que dos o más partes hacen acuerdos para afectar a un tercero en un proceso de contratación estatal (Galvis-Quintero, 2016). La Superintendencia de Industria y Comercio se refiere a la colusión como “una de las principales vías de defraudación del Estado” (2013), puesto que en contratación pública suele provocar que la cuantía de los procesos incremente. Por este motivo, analizar cómo se desarrollan dichas prácticas puede aportar a los esfuerzos anticorrupción y a salvaguardar los recursos públicos. Aun así, la gran cantidad de datos no estructurados que se encuentran disponibles dificulta el trabajo de los organismos de control, por lo cual es necesario aumentar la capacidad de análisis y tener una mejor respuesta. Ante esto el Machine Learning se convierte en una herramienta útil para acelerar estos procesos pues pueden encontrar patrones complejos en un gran número de datos. Algunos ejemplos de lo anterior son las pruebas que han realizado los gobiernos de México y Ucrania para identificar casos de fraude con la implementación de algoritmos de Machine Learning.

### **3. Objetivos**

- **Objetivo General:**

Realizar un análisis de las relaciones entre los proponentes de contratos de vías primarias en Colombia, dirigido a facilitar la identificación de posibles casos que afecten la competitividad de dichos procesos.

- **Objetivos específicos:**

- Construir una base datos con la información de los proponentes que participaron en los procesos de contratación de vías primarias entre los años 2014 y 2018 a nivel nacional.
- Realizar una revisión de literatura para la caracterización de los proponentes y su trayectoria.
- Seleccionar un algoritmo de aprendizaje no supervisado a partir de la evaluación de un conjunto de métricas.
- Identificar relaciones relevantes entre los proponentes que puedan afectar la competitividad de los procesos de contratación mediante la implementación del algoritmo.
- Desarrollar un sistema de alertas que permita establecer la probabilidad de que, dadas sus características, un proponente pueda estar relacionado en prácticas que afecten la competitividad de los procesos de contratación.

### **4. Marco teórico**

- **Contexto**

La compra pública es el mecanismo con el que las entidades gubernamentales adquieren los productos y servicios que requieren para cumplir con sus funciones. En Colombia ésta se realiza a través del Sistema de Compra Pública SECOP, plataforma que media estas transacciones. El gasto en compra pública representa alrededor del 15% del PIB del país, uno de los más altos en la región según la OCDE (2014). Por este motivo, analizar los factores que

influyen en el sistema de compra pública puede ayudar a identificar ineficiencias que permiten que se dé el derroche de recursos públicos.

En el país, las modalidades de contratación en el régimen general de acuerdo con el artículo 2 de la ley 1150 de 2007 son licitación pública, selección abreviada, concurso de méritos y contratación directa. En esta misma ley, como mecanismos de la modalidad de selección abreviada, se establecieron la subasta inversa y la contratación de mínima cuantía. Con esta última se buscó agilizar los procesos de compras cuyo valor no superara el 10% de la mínima cuantía establecida para la entidad. Por otro lado, existe el régimen especial para algunas entidades estatales que cuentan con la facultad de aplicar reglas distintas a las previstas en las Leyes 80 de 1993 y 1150 de 2007, además de contar con su propio manual (Colombia Compra Eficiente, 2016).

Al observar el valor total contratado según la modalidad para el periodo 2014 – 2018, se encuentra que el régimen especial concentra el 34% del monto total de los contratos, la contratación directa el 24%, la licitación pública el 22% y las Asociaciones Público-Privadas el 5%. Se debe tener en cuenta que bajo la modalidad de Contratación Directa se realizan los contratos de prestación de servicios; el número de procesos en esta modalidad supera los dos millones, lo que se refleja en la participación sobre el valor total contratado. Por otro lado, en la modalidad de Licitación Pública el número de procesos es 36.753, los cuales se caracterizan por tener grandes cuantías lo que explica el porcentaje que representa sobre el valor total contratado. De forma similar, las Asociaciones Público-Privadas se caracterizan por el alto valor contratado, este 5% corresponde a solo 67 procesos, como se puede ver en el gráfico 2.

## MODALIDAD DE CONTRATACIÓN SEGÚN PORCENTAJE DEL VALOR TOTAL CONTRATADO

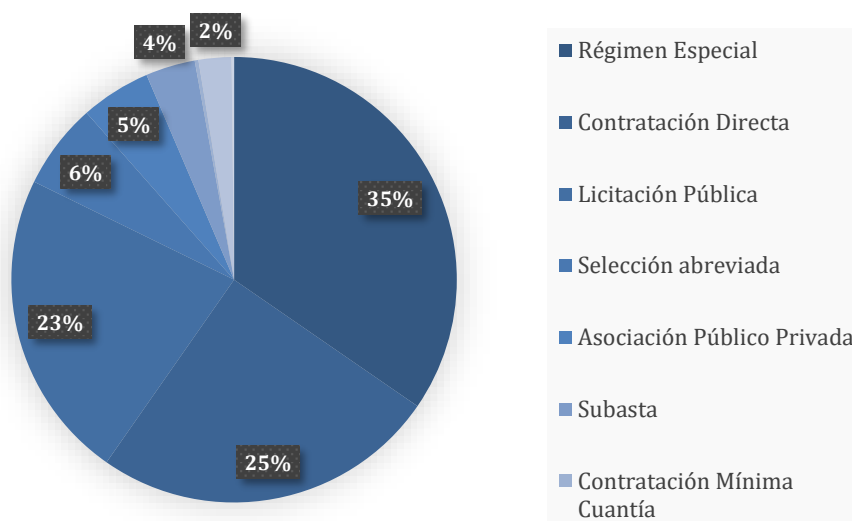


Gráfico 2. Fuente: Elaboración propia con base en SECOP I

En la modalidad de Licitación Pública se destacan los procesos asociados a obra pública, los cuales para el periodo 2014 – 2018 superaron los setecientos billones de pesos; en estos se encuentran gran parte de las apuestas de infraestructura del país. Dada la importancia que tiene para el desarrollo del país, este análisis se enfocará en los contratos de vías primarias, las cuales son definidas por el Instituto Nacional de Vías (INVÍAS) como:

*“aquellas troncales, transversales y accesos a capitales de Departamento que cumplen la función básica de integración de las principales zonas de producción y consumo del país y de éste con los demás países. Este tipo de carreteras pueden ser de calzadas divididas según las exigencias particulares del proyecto, [...] deben funcionar pavimentadas.” (INVÍAS, 2016)*

En el gráfico 3 se puede observar las vías primarias y las entidades estatales encargadas de su administración, principalmente INVÍAS y la Agencia Nacional de Infraestructura (ANI). La red de vías primarias tiene una extensión aproximada de 19.300 km, de los cuales 7.000 se encuentran bajo la administración de INVÍAS. Un informe de esta entidad con corte al 28 de diciembre de 2018 indica que solo 13,61% se encuentra en excelentes condiciones, lo cual se atribuye en buena parte a la ineficiencia en la ejecución de los proyectos y los casos de corrupción que se han dado en los últimos años (Bohórquez, 2019).

## VÍAS PRIMARIAS

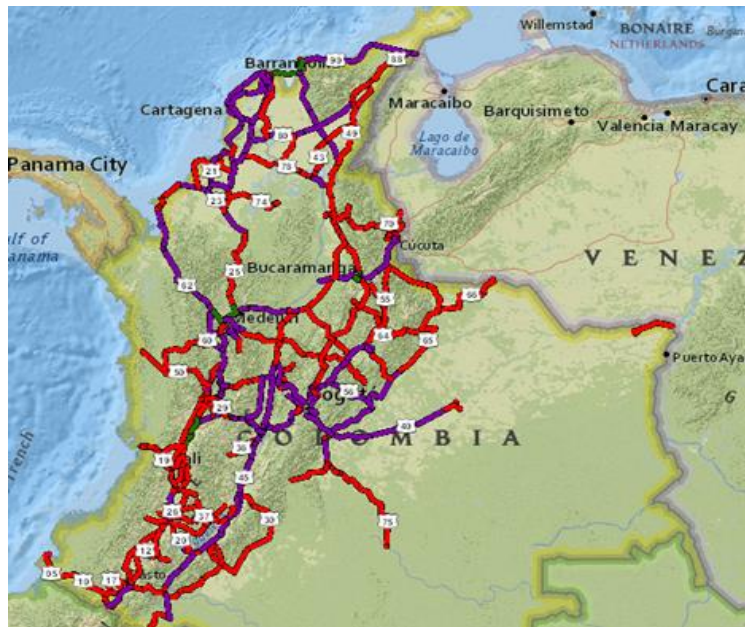


Gráfico 3. Fuente: Mapa de carreteras INVÍAS <https://hermes.invias.gov.co/carreteras/>

La eficiencia de los procesos de compra pública es fundamental para asegurar la materialización de la política pública en la forma de bienes y servicios para los ciudadanos, puesto que el acceso se ve afectado directamente cuando los recursos son derrochados. De forma general, el derroche de recursos se puede considerar activo o pasivo (Bandiera , Prat, & Valletti, 2008) de acuerdo con el contexto en que ocurra. Por una parte, se considera pasivo cuando el derroche ocurre por ineficiencias en el proceso, sin estar sujeto a intereses personales de los actores involucrados. Así mismo, se considera activo cuando sucede debido a intereses personales de los servidores públicos, es decir, una forma de corrupción. Lo anterior afecta el desarrollo de los países al obstaculizar el crecimiento y la mejora del desempeño de los procesos socioeconómicos, siendo la corrupción la principal fuente de estas dificultades.

Según Transparencia por Colombia (2019) la percepción de corrupción para el año 2018 fue de 36 puntos sobre 100, medida que considera la percepción de la ciudadanía respecto a la gestión de recursos públicos y a la contratación estatal, teniendo el mismo puntaje que Panamá y superando en Suramérica únicamente a Ecuador (34) y Venezuela (18). En Colombia, el sistema de compra pública es la estructura que vincula las entidades estatales,

los actores del mercado, las autoridades y los sistemas de información para la toma de decisiones acerca del gasto público.

Se han realizado diversos esfuerzos para enfrentar la corrupción. Dentro de estos se encuentra la metodología del Instituto Mexicano para la Competitividad (IMCO) diseñada para evaluar el riesgo en el sistema de compra pública y que se basa en tres dimensiones de análisis - falta de competencia, falta de transparencia y violaciones en los procesos - que consideran 43 indicadores en total (Zuleta, Ospina, & Caro, 2019).

Este trabajo se enfocará en analizar las prácticas que restringen la competencia, especialmente las que se encuentran asociadas a colusión. De acuerdo con su sentencia del 26 de enero de 1995, la Corte Suprema de Justicia, Sala Civil, precisó que “las maniobras fraudulentas pueden provenir del acuerdo de las partes (colusión) para perjudicar a terceros, o de una de las partes para perjudicar a la otra”. Es decir, la colusión hace referencia a la acción de pactar contra un tercero en compra pública, que afecta tanto a los oferentes que no hicieron parte del acuerdo como al mercado. Estas prácticas provocan un incremento en el precio de los bienes y servicios contratados por el Estado; según la OCDE, el sobreprecio frente a un mercado competitivo se encuentra entre el 10% y 20%.

- **Revisión bibliográfica**

El marco normativo del sistema de compra pública se encuentra representado por un amplio número de leyes y decretos, dentro de los cuales se encuentran:

#### **MARCO NORMATIVO DEL SISTEMA DE COMPRA PÚBLICA**

<b>Norma</b>	<b>Objeto</b>
Ley 80 de 1993	Estatuto General de Contratación de la Administración Pública
Ley 1150 de 2007	Por medio de la cual se introducen medidas para la eficiencia y la transparencia en la Ley 80 de 1993 y se dictan otras disposiciones generales sobre la contratación con Recursos Públicos.
Ley 1474 de 2011	Por la cual se dictan normas orientadas a fortalecer los mecanismos de prevención, investigación y sanción de actos de corrupción y la efectividad del control de la gestión pública.



Ley 1712 de 2014	Por medio de la cual se crea la Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional y se dictan otras disposiciones.
Decreto reglamentario 1510 de 2013	Por el cual se reglamenta el sistema de compras y contratación pública
Decreto 1081 de 2015	Por medio del cual se expide el Decreto Reglamentario Único del Sector Presidencia de la República.

Adicionalmente, la Superintendencia de Industria y Comercio (SIC) fue designada como Autoridad Nacional de Protección de la Competencia por la Ley 1340 de 2009 para velar por el cumplimiento de las disposiciones en este ámbito. Con el fin de facilitar la vigilancia de una sana competencia en los procesos de Licitación Pública, la SIC desarrolló una guía práctica para combatir la colusión. En dicha guía enuncia algunas de las prácticas asociadas a colusión dentro de un proceso de Licitación Pública, así como señales de advertencia que pueden ayudar a identificar estas conductas (SIC, 2013).

Algunas de las formas que según la SIC puede tomar la colusión son, por ejemplo, el intercambio de información entre los funcionarios y proponentes para direccionar la adjudicación; rotación de ofertas, es decir, cuando los proponentes coludidos acuerdan turnos para ganar la licitación; asignación de mercado, donde los proponentes se reparten el mercado o zonas geográficas; entre otras. Por otro lado, dentro de las señales de advertencia se encuentra que un proponente tenga frecuentemente la mejor oferta; algunos proponentes desisten inesperadamente de participar en el proceso; el proponente que desiste del contrato posteriormente se encuentra como subcontratista; hay diferencias significativas entre el precio de la oferta ganadora y las demás que fueron presentadas; los proponentes han trabajado juntos en uniones temporales o consorcios para participar en otras licitaciones; entre otras. Lo anterior ofrece un punto de partida para identificar los comportamientos que pueden dar señales de alerta ante los posibles casos de colusión en procesos de contratación.

A nivel internacional, el Banco Mundial ha realizado esfuerzos para detectar y prevenir contratos de desarrollo en los que se puedan dar casos de fraude, corrupción o colusión. Lo anterior es de gran importancia teniendo en cuenta que, según las estimaciones del Foro

Económico Mundial, el costo de dichas prácticas se encuentra alrededor del 5% del producto interno global, es decir, cerca de 2.6 trillones de dólares al año (Irisova, 2014). En una de estas iniciativas el Banco Mundial, junto con investigadores de varias universidades en Estados Unidos, implementaron algoritmos de Machine Learning para estimar el riesgo asociado a contratos de desarrollo internacional y de esta forma poder priorizar cuales deben ser investigados. Para este proceso se utilizó una base de datos con cerca de 200.000 contratos adjudicados por el Banco Mundial desde el año 2000, de los cuales alrededor del 2% fueron investigados por la vicepresidencia de integridad (INT), siendo este porcentaje la variable de respuesta usada para entrenar el modelo. Para la construcción de los modelos incluyeron variables asociadas al contrato y características del contratista que reflejaban su experiencia en el sector, por ejemplo, el porcentaje de contratos realizados en los últimos 5 años relacionados con el sector. Compararon el desempeño de varios algoritmos tales como Random Forest, Suport Vector Machine, Boosting y AdaBoost. Este proceso les permitió incrementar en un 84% la tasa de éxito de las investigaciones que estaban vigentes y hacer el proceso más eficiente pues les facilita a los investigadores priorizar los contratos que deben ser analizados (Grace et. al., 2016).

Las herramientas de Machine Learning permiten establecer qué variables son significativas al momento de caracterizar el riesgo asociado a un contrato para de esta forma predecir cuáles de ellos pueden estar involucrados en fraude, corrupción o colusión. Sin embargo, en Colombia la implementación de este modelo se dificulta debido a la limitada cantidad de información disponible sobre contratos en los que se han dado sanciones por corrupción y, aún en menor proporción, por colusión. Lo anterior impide el uso de algoritmos supervisados que requieren datos “etiquetados” en los que se evidencie el comportamiento de interés.

Por otro lado, se encontró un estudio que utiliza mapas autoorganizados - un tipo de red neuronal - para identificar posibles casos de corrupción en provincias de España (López-Iturriaga y Sanz, 2017). En dicho estudio se usó información sobre las condiciones económicas y políticas de las 52 provincias de España con el fin de encontrar patrones y crear perfiles de riesgo que faciliten detectar corrupción en cada una de ellas. Lo anterior mostró que las provincias con factores políticos y sociales, tales como mayores impuestos y partidos políticos que conservaban cargos por más tiempo, al parecer presentaban más relación con casos de

corrupción pública. De forma similar Issa y Vaserhelyi (2011), mediante la implementación del algoritmo K-means para detectar anomalías en actividades empresariales, muestran cómo los algoritmos de segmentación pueden ayudar a identificar procesos fraudulentos. Para esto utilizaron bases de datos de una compañía de telecomunicaciones acerca de procesos de reembolso y por medio de la implementación del algoritmo de clustering identificaron transacciones que por sus características podrían ser fraude. Los algoritmos mencionados anteriormente se encuentran dentro de los métodos de aprendizaje no supervisado, un área de Machine Learning que permite encontrar patrones dentro de un grupo de datos.

## **5. Metodología**

- **Datos del sistema electrónico de compra pública**

Con el fin de comprender cómo se desarrollan las prácticas restrictivas de la competencia en los procesos de contratación de vías primarias, se analizó la información disponible en el sistema de información SECOP I. La información principal de este sistema se encuentra recopilada en la plataforma *Datos Abiertos*, en la que se puede consultar acerca de las entidades contratantes, los proponentes y detalles de cada contrato como el objeto, modalidad de contratación, valor total del contrato, plazo de ejecución, adiciones, entre otros.

De acuerdo con la Ley 1150 de 2007, la Ley 1712 de 2014 y el Decreto 1510 de 2013, las entidades deben publicar toda la información relativa a la gestión y ejecución del contrato (desde la etapa precontractual hasta la liquidación) en las plataformas del SECOP. Por este motivo, se cuenta con los documentos que se generan en el proceso tales como adendas, actas de adjudicación, contratos, modificaciones, entre otros, lo cual da la posibilidad de analizar en detalle los procesos para identificar posibles irregularidades. Con este análisis se buscó entender cómo se dan las prácticas restrictivas de la competencia dentro de los procesos de contratación de vías primarias.

Inicialmente se hizo un filtro para determinar los contratos que serían analizados, buscando así las Asociaciones Público-Privadas y licitaciones cuyo objeto de contrato estaba directamente relacionado con la construcción de vías primarias. De esta forma, se

seleccionaron 65 contratos que fueron adjudicados entre 2014 y 2018, convocados por 5 entidades como se puede ver en el siguiente gráfico:

#### PORCENTAJE DE CONTRATOS ANALIZADOS SEGÚN ENTIDAD

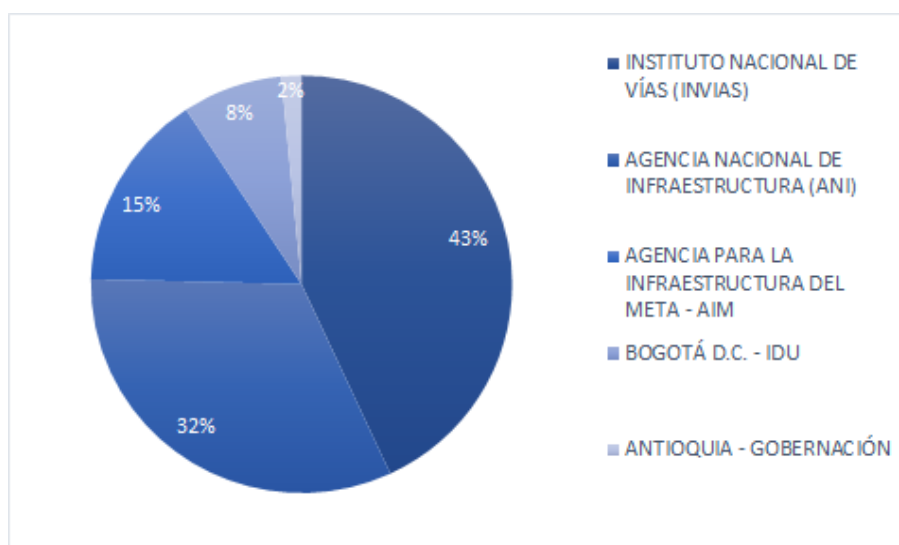


Gráfico 4. Fuente: elaboración propia.

De esta forma, el 43% de los contratos analizados estaban a cargo del Instituto Nacional de Vías (INVIAS), el 32% de la Agencia Nacional de Infraestructura (ANI), un 15% de la Agencia para la Infraestructura del Meta (AIM), por último, el IDU en Bogotá y la Gobernación de Antioquia con un 8% y 2%, respectivamente.

Los departamentos donde se ejecutaron las obras de estos contratos se pueden ver en las siguientes gráficas, en las que se encuentra el número y el valor total contratado. Es importante tener en cuenta que, por la naturaleza de las vías primarias, los contratos pueden ejecutarse en más de un departamento; sin embargo, para la elaboración de estos mapas solo se tuvo en cuenta el departamento que registra el SECOP I.

## DISTRIBUCIÓN GEOGRÁFICA DE CONTRATOS SEGÚN DE ACUERDO CON LA CANTIDAD ANALIZADA Y VALOR TOTAL CONTRATADO

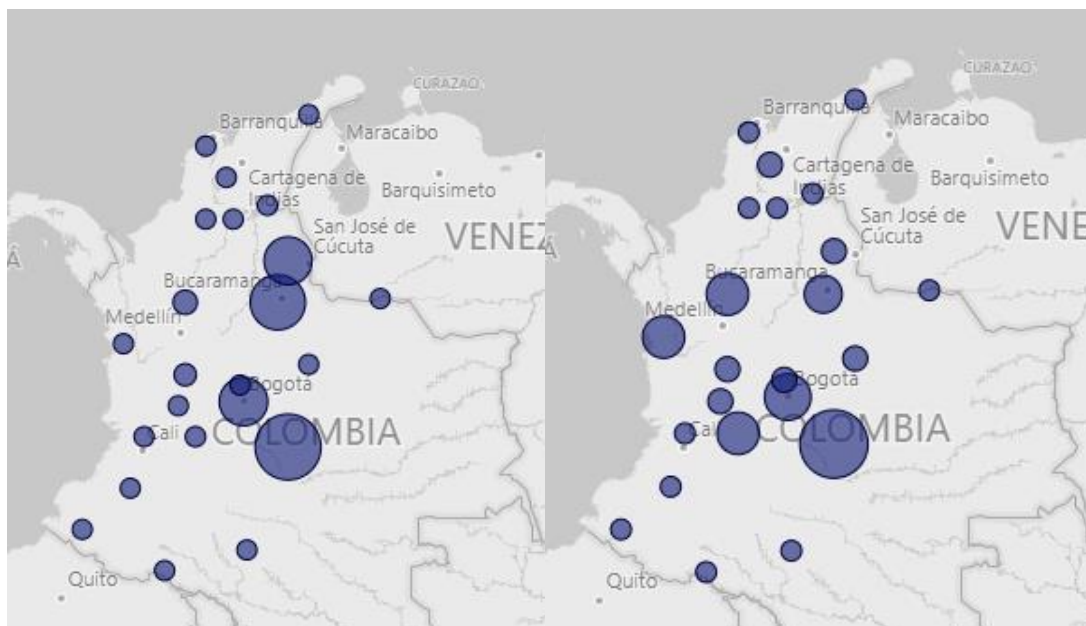


Gráfico 5. a) Número de contratos | b) Valor contratado Fuente: elaboración propia

Para cada uno de los contratos se revisó el proceso contratación y, verificando las actas de adjudicación, se identificó los proponentes que participaron en el proceso, los cuales fueron en total 1.039. Con un máximo de 60 y un mínimo de 1, se encontró que en la mayoría de los contratos se presentaron entre 3 y 27 proponentes. No obstante, para el 50% de los contratos analizados se presentaron 10 o menos proponentes, mientras que la media es 15.98 proponentes, como se aprecia en la gráfica siguiente. Esto indica que los datos son asimétricos y se concentran en valores inferiores a 10.

## DISTRIBUCIÓN DEL NÚMERO DE PROPONENTES PLURALES POR CONTRATO

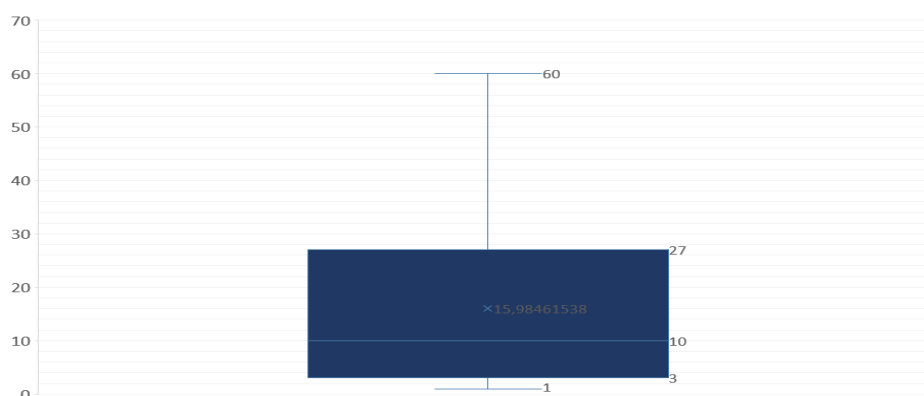


Gráfico 6. Fuente: elaboración propia.

Sin embargo, debido a las dimensiones de los contratos de obra pública, gran parte de estos proponentes son plurales, es decir son consorcios, uniones temporales o promesas de sociedad futura, por lo que están conformados por más de una empresa y/o personas naturales. Por este motivo se revisó para cada uno de los proponentes la forma como estaban constituidos. Este procedimiento permitió reconocer 583 participantes entre personas jurídicas y naturales.

Con esta información y haciendo uso de la plataforma del Registro Único Empresarial (RUES), la cual consolida la información de todas las Cámaras de Comercio del país, se buscó para cada una de las personas jurídicas y naturales la fecha en que fue matriculada, el número de empleados y las actividades económicas registradas. Lo anterior, como para tener variables que muestren de forma aproximada la experiencia, tamaño y la relación que tienen las actividades de los oferentes con los contratos de obra pública.

Durante el proceso de búsqueda de información se detectaron 39 personas jurídicas y naturales que no se encontraban en el RUES. Al revisar estos casos, se encontró que algunos corresponden a personas jurídicas que han sido sancionadas y liquidadas por diversos motivos, como incumplimientos a contratos e incluso casos de corrupción. Sin embargo, para algunas de estas personas jurídicas y naturales no se encontró información tanto en los registros públicos, como en una búsqueda más amplia; para estos casos, a las variables de experiencia, empleados y el puntaje se les asignó 0 como valor.

A continuación se presenta el histograma de frecuencias para los años de experiencia, en el que se observa que un alto número de personas jurídicas y naturales cuenta con menos de 12 años de experiencia, mientras que un reducido número de ellas tiene más de 48 años de experiencia.

### HISTOGRAMA DE FRECUENCIAS PARA LOS AÑOS DE EXPERIENCIA

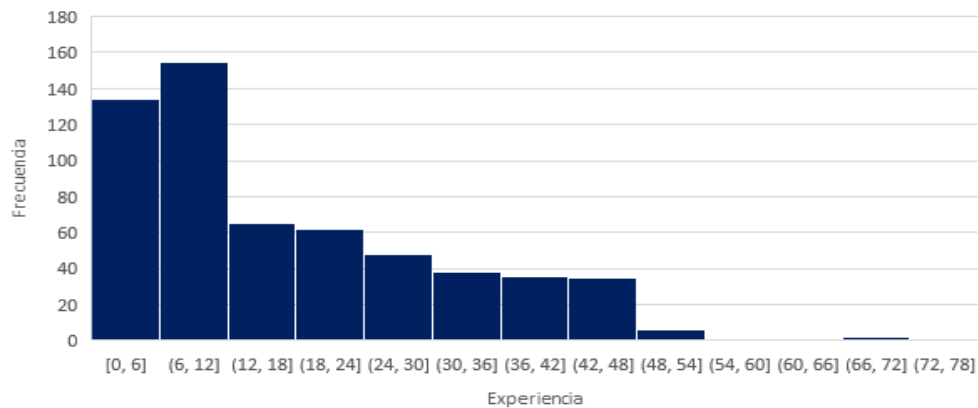


Gráfico 7. Fuente: elaboración propia.

Respecto al número de empleados se encuentra que la mayoría registran ante el RUES 40 o menos empleados, tal como se puede observar en el siguiente histograma:

### HISTOGRAMA DE FRECUENCIAS PARA EL NÚMERO DE EMPLEADOS

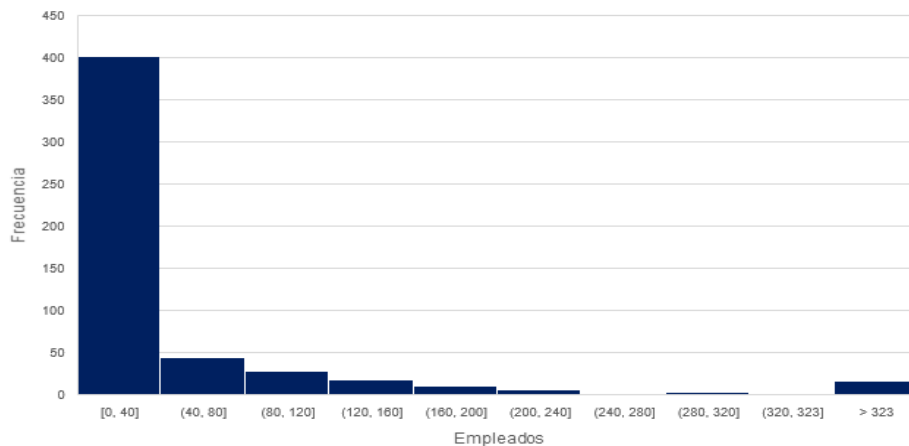


Gráfico 8. Fuente: elaboración propia.

Respecto a la relación que tenían las actividades económicas que registran ante la Cámara de Comercio con los contratos de obra pública, se estableció un puntaje para cada actividad de la siguiente forma:

$$puntuaje = \begin{cases} 2 & \text{alta relación} \\ 1 & \text{alguna relación} \\ 0 & \text{no tiene relación} \end{cases}$$

Con esta información, para cada uno de las personas jurídicas y naturales se promediaron los puntajes de sus actividades económicas. El histograma de frecuencia permite ver que la gran mayoría tienen una alta relación con el objeto de los contratos.

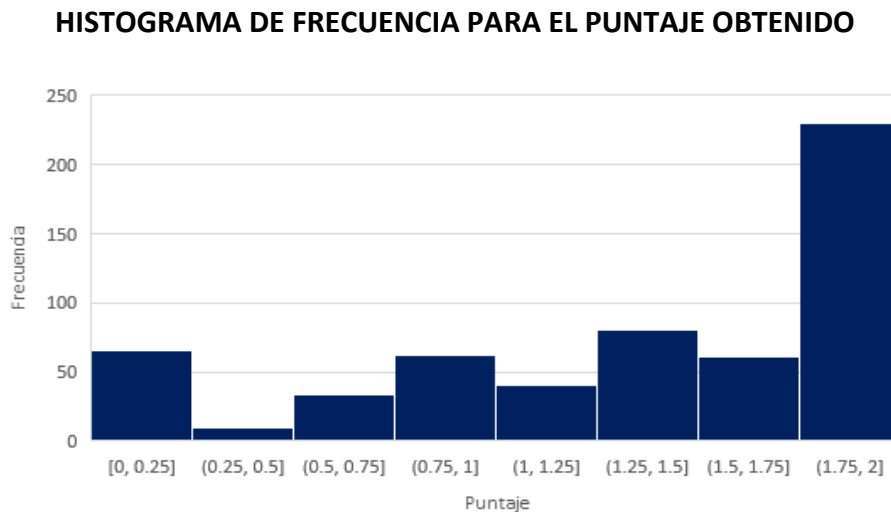


Gráfico 9. Fuente: elaboración propia.

Las variables que se utilizaron en los modelos fueron años de experiencia, número de empleados y el puntaje de las actividades económicas de cada oferente. Por otra parte, para cada uno de los 65 contratos analizados se consideró si la persona jurídica o natural participó o no, y en el caso de que participara, se tuvo en cuenta si lo hizo de forma individual o junto con otros oferentes.

- **Modelos de segmentación**

Se utilizaron algoritmos de clustering o segmentación, técnicas de aprendizaje no supervisado con las que se busca organizar  $N$  objetos en  $K$  grupos o segmentos de acuerdo con su semejanza (Hastie, Tibshirani and Friedman, 2009). Según la clasificación de los autores, se consideraron 4 clases de algoritmos que se exponen a continuación:

**Método jerárquico:**

Este algoritmo busca que las observaciones al interior de cada segmento sean lo más semejantes entre sí. Para esto, inicialmente se considera a cada una de las observaciones como un segmento y a continuación se unen las más cercanas para formar un nuevo segmento. El proceso se repite hasta que todas las observaciones se hayan agrupado. Este



método permite la construcción de árboles o dendrogramas que pueden cortarse en un punto específico para obtener un número deseado de segmentos.

### **K-means:**

Es un método iterativo que busca minimizar la varianza al interior de cada segmento. Para esto, el algoritmo asigna cada observación al segmento que tiene el promedio más cercano. De forma que, inicialmente, las observaciones tienen una asignación arbitraria y posteriormente cada observación se reasigna de acuerdo con el siguiente criterio:

$$C(i) = \underset{1 \leq k \leq K}{\operatorname{argmin}} ||x_i - m_k||^2$$

donde  $C(i)$  representa el segmento al que fue asignada la observación  $i$  y  $m_k$  el promedio del segmento  $k$ . Después de cada iteración se actualizan los promedios de cada segmento; el algoritmo se detiene cuando los centros dejan de cambiar.

### **K-medoids:**

También conocido como Partición Alrededor de Mediods (PAM), es un algoritmo iterativo similar a K-means aunque se considera más robusto, el cual intenta minimizar la distancia entre las observaciones. Sin embargo, a diferencia de K-means, utiliza el centroide del segmento para calcular las distancias al interior de este.

### **SOM:**

Un mapa autoorganizado o SOM (por sus siglas en inglés) es un tipo de red neuronal que permite representar observaciones de altas dimensiones en un sistema de coordenadas bidimensional al que también se le conoce como un mapa topológico restringido. Dicho mapa está formado por una rejilla de nodos que pueden ser de forma rectangular o hexagonal, donde el algoritmo asocia a cada nodo un vector de pesos con el mismo tamaño de las variables de entrada. Se busca reducir la distancia entre los vectores asociados a los nodos y los vectores de entrada. Una vez se realiza este proceso se pueden clasificar las observaciones de acuerdo con su cercanía a los nodos (Hollmen, 1996).

- **Validación y selección del modelo de segmentación**

De acuerdo con lo propuesto por Brock et al. (2008), para evaluar algoritmos de segmentación se utilizan las siguientes métricas internas y de estabilidad para verificar cuál algoritmo tiene mejor desempeño.

### Métricas internas

Estas medidas buscan que las observaciones al interior de un segmento sean muy similares entre ellas y muy diferentes entre segmentos. Se consideraron las siguientes métricas:

#### **Conectividad**

Para una segmentación particular de un conjunto de  $N$  observaciones, la conectividad se calcula como:

$$conectividad = \sum_{i=1}^N \sum_{j=1}^L x_{i,nn_{i(j)}}$$

donde  $nn_{i(j)}$  toma el valor de 0 cuando el  $j$ -ésimo vecino de la observación  $i$  se encuentra en el mismo segmento o  $1/j$  de lo contrario, y  $L$  es un parámetro que determina el número de vecinos que contribuyen a la medida. Al realizar el cálculo de esta métrica para los cuatro algoritmos en consideración se encontró el siguiente resultado:

#### **COMPARACIÓN DE CONECTIVIDAD POR MÉTODO DE SEGMENTACIÓN**

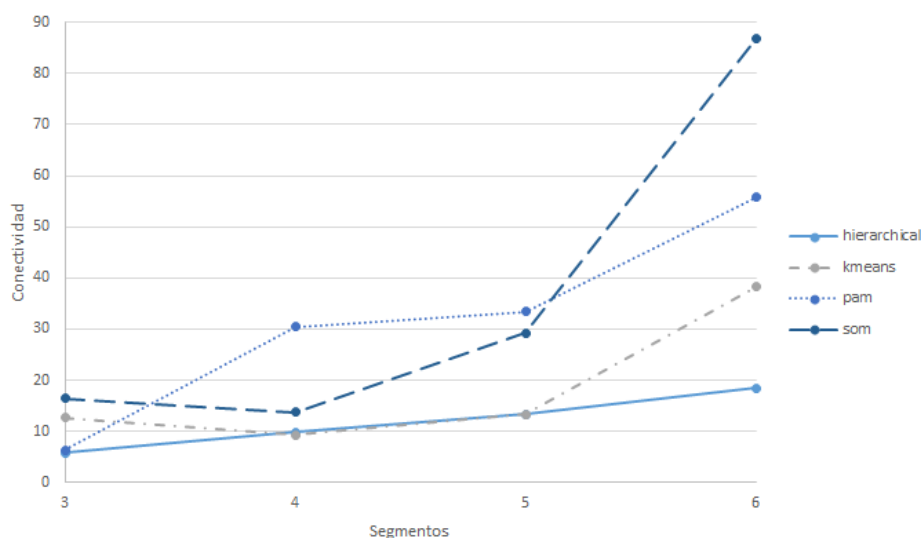


Gráfico 10. Fuente: elaboración propia.

Dado que se busca minimizar el valor de conectividad, el método jerárquico tiene el mejor desempeño.

### Índice de Dunn

Es la razón entre la mínima distancia  $d(i, j)$  entre las observaciones que no se encuentran en el mismo segmento y la máxima distancia  $d'(i, j)$  de las observaciones al interior de los segmentos, de la siguiente forma:

$$Dunn = \frac{\min_{1 \leq i < j \leq n} d(i, j)}{\max_{1 \leq i < j \leq n} d'(i, j)}$$

Esta métrica se evaluó para los cuatro algoritmos en consideración y se encontró el siguiente resultado:

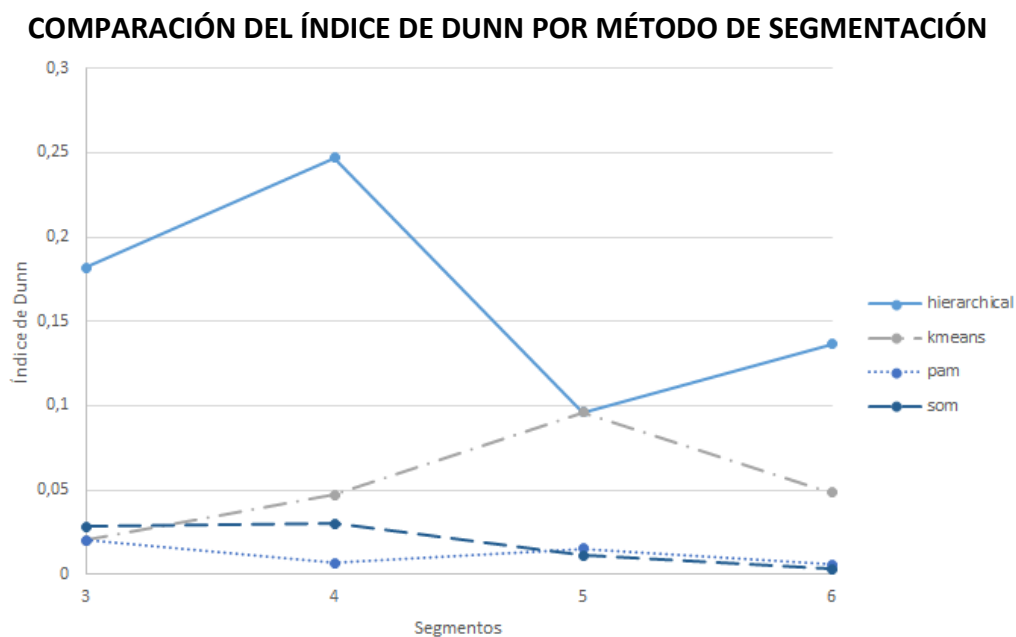


Gráfico 11. Fuente: elaboración propia.

Dado que se busca maximizar el valor del índice de Dunn, el método jerárquico tiene el mejor desempeño.

### Ancho de silueta

Corresponde al promedio del valor de silueta de cada observación en la muestra. El valor de silueta representa la semejanza entre las observaciones asignadas a un mismo segmento. Para una observación  $i$  el valor de silueta se calcula como:

$$S_i = \frac{b_i - a_i}{\max(b_i, a_i)}$$

donde  $a_i$  representa la distancia promedio entre la observación  $i$  y las demás observaciones del mismo segmento  $C(i)$ , y  $b_i$  representa la distancia promedio entre la observación  $i$  y las observaciones del segmento más cercano  $C(k)$

$$a_i = \frac{1}{n_{C(i)}} \sum_{j \in C(i)} \text{dist}(i, j)$$

$$b_i = \min \sum_{j \in C(k)} \frac{\text{dist}(i, j)}{n_{C(k)}}$$

Esta métrica se evaluó para los cuatro algoritmos en consideración y se encontró el siguiente resultado:

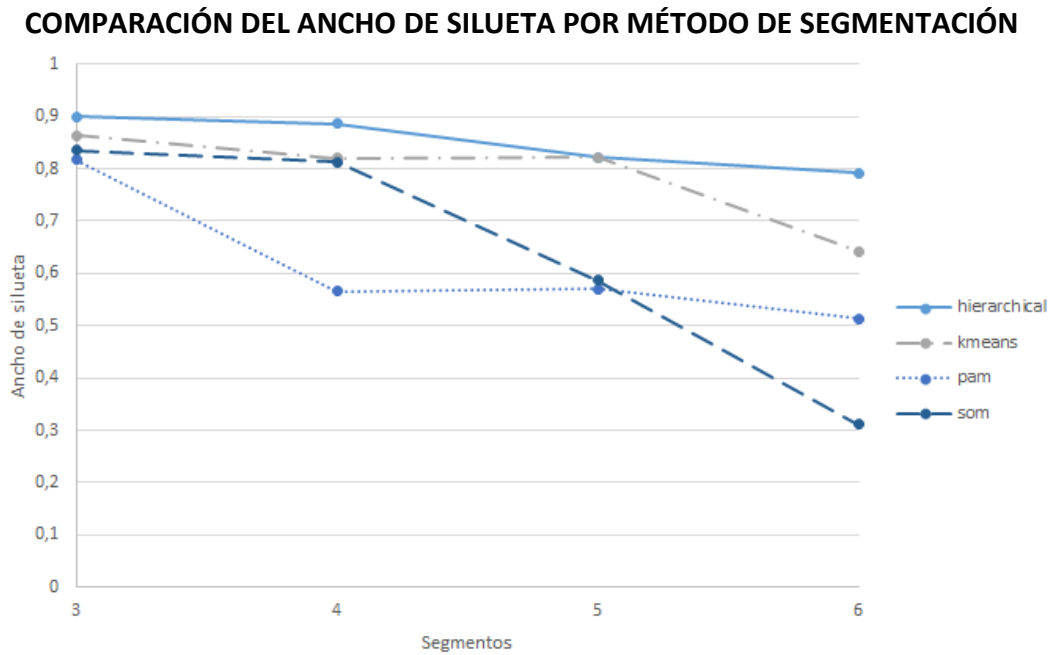


Gráfico 12. Fuente: elaboración propia.

Dado que se busca maximizar el valor del ancho de silueta, el método jerárquico tiene el mejor desempeño.

### Métricas de estabilidad

Estas medidas comparan los resultados obtenidos al hacer la segmentación con toda la información disponible frente a los que se obtienen al remover un variable a la vez.

Denotando  $N$  como el conjunto de observaciones y  $M$  como el conjunto de variables, se consideraron las siguientes métricas:

## APN

La proporción promedio de no superposición o APN (por sus siglas en inglés) considera la proporción promedio de observaciones que no se encuentran en el mismo segmento al eliminar una variable del conjunto de datos. Se calcula de la siguiente forma:

$$APN = \frac{1}{MN} \sum_{i=1}^N \sum_{\ell=1}^M \left( 1 - \frac{n(C^{i,\ell} \cap C^{i,0})}{n(C^{i,0})} \right)$$

donde  $C^{i,0}$  representa el segmento que contiene la observación  $i$  en la configuración original y  $C^{i,\ell}$  representa el segmento que contiene la observación  $i$  al remover la variable  $\ell$ . Esta métrica se evaluó para los cuatro algoritmos en consideración y se encontró el siguiente resultado:

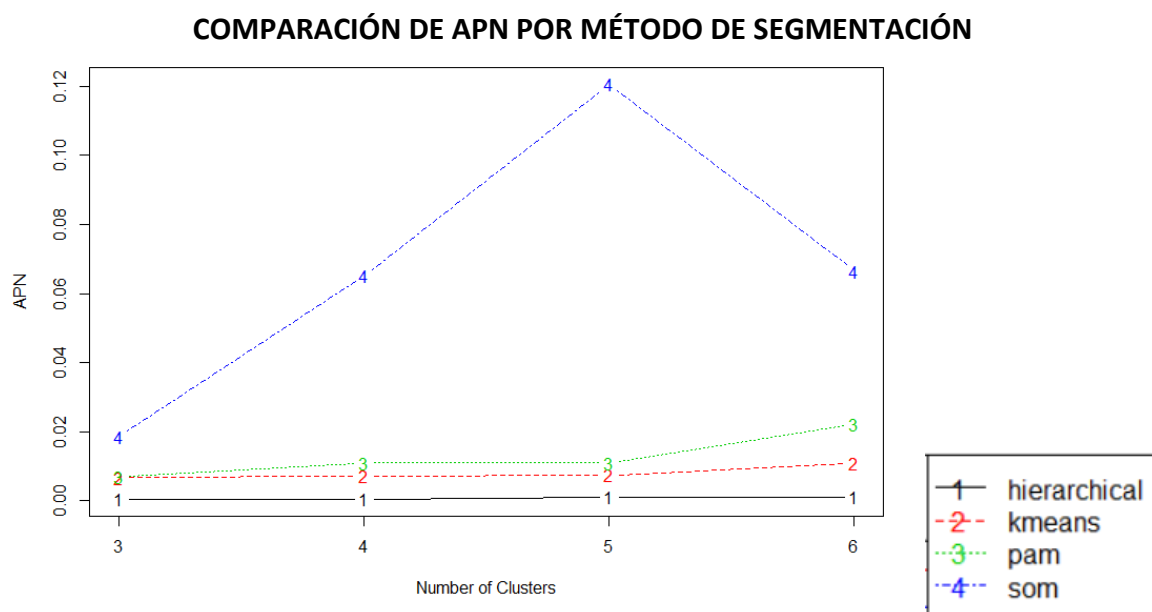


Gráfico 13. Fuente: elaboración propia.

Dado que se busca minimizar el valor del APN, el método jerárquico tiene el mejor desempeño.

## AD

La distancia promedio o AD (por sus siglas en inglés) representa la distancia promedio entre las observaciones asignadas a un mismo segmento usando toda la información y al remover una columna a la vez, de la siguiente forma:

$$AD = \frac{1}{MN} \sum_{i=1}^N \sum_{\ell=1}^M \frac{1}{n(C^{i,0})n(C^{i,\ell})} \left[ \sum_{i \in C^{i,0}, j \in C^{i,\ell}} dist(i, j) \right]$$

Esta métrica se evaluó para los cuatro algoritmos en consideración se encontró el siguiente resultado:

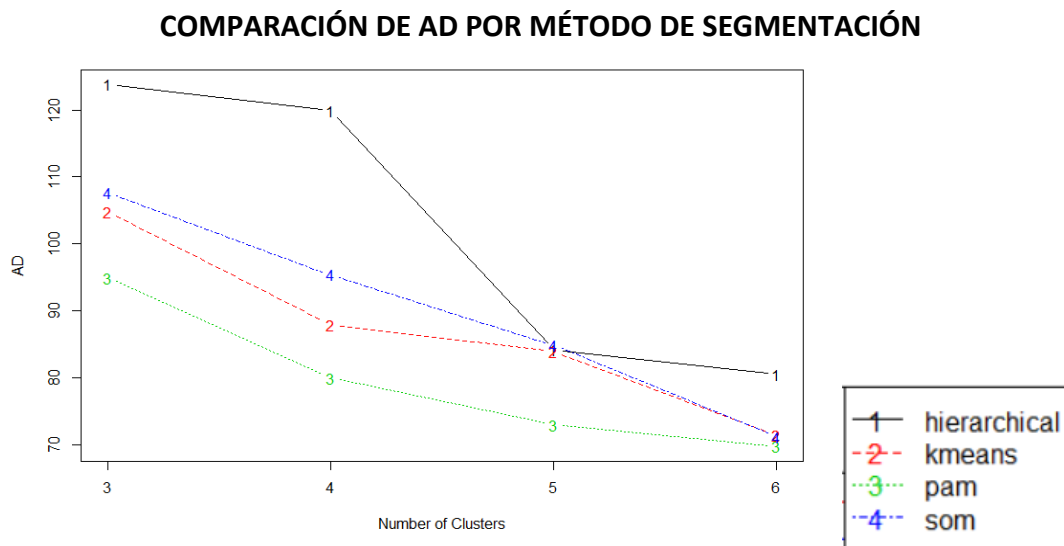


Gráfico 14. Fuente: elaboración propia.

Dado que se busca minimizar el valor del AD, el método pam o k-medioides tiene el mejor desempeño.

## ADM

La distancia promedio entre las medias o ADM (por sus siglas en inglés) calcula la distancia promedio entre los centros de los segmentos para las observaciones ubicadas en el mismo segmento usando toda la información y al remover una columna a la vez, de la siguiente forma:

$$ADM = \frac{1}{MN} \sum_{i=1}^N \sum_{\ell=1}^M dist(\bar{x}_{C^{i,0}}, \bar{x}_{C^{i,\ell}})$$

donde  $\bar{x}_{C^{i,0}}$  es el promedio de las observaciones en el segmento que contiene la observación  $i$  al utilizar toda la información y  $\bar{x}_{C^{i,\ell}}$  es el promedio de las observaciones en el segmento que contiene la observación  $i$  al remover la variable  $\ell$ . Esta métrica se evaluó para los cuatro algoritmos en consideración se encontró el siguiente resultado:

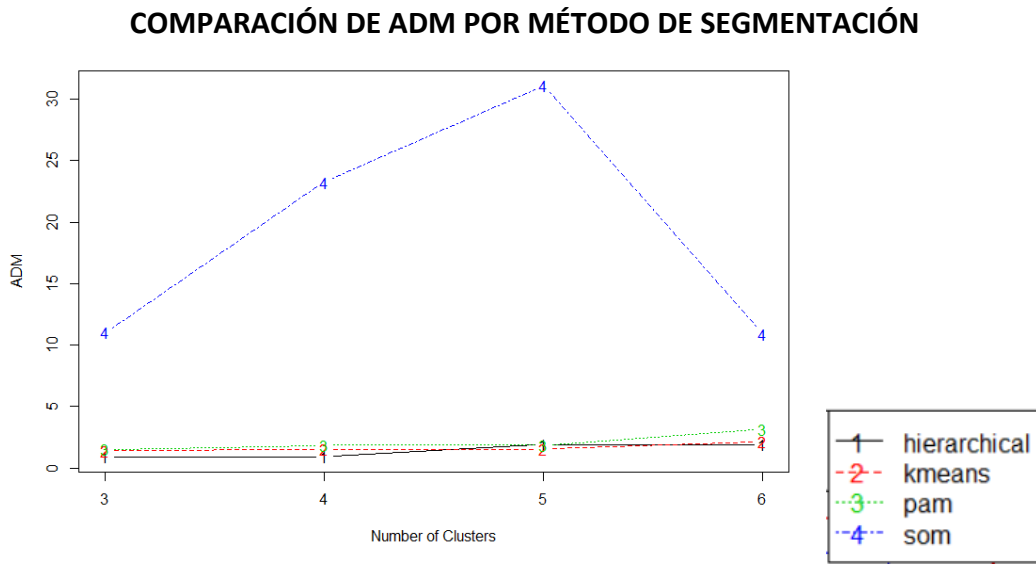


Gráfico 15. Fuente: elaboración propia.

Dado que se busca minimizar el valor del ADM, el método jerárquico tiene el mejor desempeño.

### FOM

La figura de mérito o FOM (por sus siglas en inglés) mide el promedio de la varianza de las observaciones al interior de los segmentos al eliminar una variable. Entonces:

$$FOM = \frac{1}{M} \sum_{\ell=1}^M \sqrt{\frac{1}{N} \sum_{k=1}^K \sum_{i \in C_k(\ell)} dist(x_{i,\ell}, \bar{x}_{C_k(\ell)})}$$

donde  $k$  representa el número de segmentos formados al remover la variable  $\ell$ . Esta métrica se evaluó para los cuatro algoritmos en consideración y se encontró el siguiente resultado:

### COMPARACIÓN DE APN POR MÉTODO DE SEGMENTACIÓN

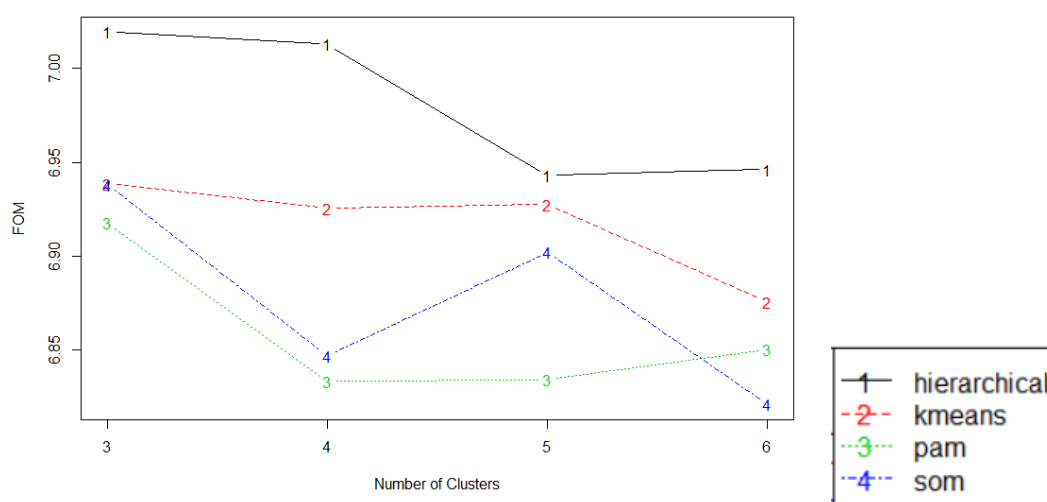


Gráfico 16. Fuente: elaboración propia.

Dado que se busca minimizar el valor del FOM, el método pam o k-medioes es el que presenta mejor desempeño.

Teniendo en cuenta los resultados de las métricas se escoge el método jerárquico para realizar la segmentación de los datos. Con la información aquí obtenida se utilizarán modelos de clasificación con las variables de años de experiencia, número de empleados y el puntaje de las actividades económicas. Esto último constituye un sistema de alertas que permite establecer la probabilidad de que, dadas sus características, un proponente pueda estar relacionado con prácticas que afecten la competitividad de los procesos de contratación. Puesto que los algoritmos de clasificación estiman la probabilidad de que una observación pertenezca a cada clase y la asigna en la que tiene mayor valor, para este caso las clases corresponden a los segmentos que se obtengan después de la implementación del método jerárquico.

- **Modelos de clasificación**

#### **Random forest**

De acuerdo con Breiman (2001) es un algoritmo de aprendizaje conjunto que construye múltiples árboles predictores no correlacionados y los combina para captar estructuras complejas de interacción entre los datos. Para esto ensambla los árboles predictores usando un vector de valores aleatorios generados para cada uno de ellos, dichos vectores son



independientes e igualmente distribuidos, lo que permite reducir la varianza y evitar el sobreajuste mejorar la precisión y estabilidad. Debido a que el algoritmo ensambla árboles individuales para cada uno de ellos se calcula la *Importancia de Gini* que es una medida que representa la importancia de las variables para estimar el resultado. Posteriormente se calcula el promedio de estos valores, denominado *Decrecimiento Promedio de Gini*, como una medida de la importancia de las variables a través de los árboles que componen el algoritmo.

### **Boosting**

Este algoritmo busca tener un alto desempeño al combinar un grupo de clasificadores “débiles” (Hastie, Tibshirani and Friedman, 2009). De acuerdo con los autores, un clasificador se considera “débil” cuando su tasa de error es ligeramente mejor que una selección aleatoria. Entonces Boosting aplica secuencialmente  $M$  algoritmos de clasificación débiles con pequeñas modificaciones para obtener un grupo de clasificadores débiles  $G_m(x), m = 1, \dots, M$ , los cuales se combinan para obtener la predicción final:

$$G(x) = I \left( \sum_{m=1}^M \alpha_m G_m(x) \right)$$

Donde  $\alpha_1, \dots, \alpha_M$  corresponden a la contribución de cada clasificador, estos valores son determinados por el algoritmo donde los clasificadores más precisos tienen una mayor participación. Al implementar Boosting se puede observar la influencia relativa de las variables empleadas al considerar el aporte en cada uno de los algoritmos de clasificación.

- **Selección del modelo de clasificación**

Para entrenar un algoritmo de clasificación la muestra de datos se divide en dos grupos, uno de entrenamiento y otro de prueba, este último se utiliza para evaluar el desempeño del algoritmo (Hastie, Tibshirani and Friedman, 2009). Para este caso se utilizó la siguiente métrica:

### **AUC - Área bajo la curva ROC**

Es una métrica que permite evaluar el rendimiento de un algoritmo de clasificación, para esto se calcula el área bajo la curva de la característica operativa del receptor o curva ROC, la cual contrasta la especificidad frente a la sensibilidad del modelo (Hanley and McNeil, 1982). Teniendo en cuenta que los algoritmos de clasificación estiman la probabilidad de que una observación pertenezca a una clase y estableciendo un umbral se define a cuál pertenece, por

ejemplo, se encuentra que la probabilidad obtenida de una observación para una clase en particular es de 0.7 y el umbral establecido es de 0.6, entonces la observación se asigna a dicha clase. Al realizar este proceso con los datos de prueba se puede identificar las observaciones que son asignadas correctamente y las que no. Esto permite construir una matriz de confusión con la que se obtiene la especificidad y la sensibilidad para un umbral establecido, y a su vez al variar el valor que toma el umbral, se obtiene la curva ROC. Para facilitar la comprensión de cómo se construye esta métrica se utilizará un ejemplo, dados los resultados de un algoritmo que busca clasificar un grupo de datos en dos clases positivos (p) y negativos (n) se puede construir la siguiente matriz de confusión para un umbral particular:

		<b>MATRIZ DE CONFUSIÓN</b>	
		Valor real	
		<b>p</b>	<b>n</b>
Resultado algoritmo	<b>p</b>	Verdadero positivo (VP)	Falso positivo (FP)
	<b>n</b>	Falso negativo (FN)	Verdadero negativo (VN)

Gráfico 17. Fuente: elaboración propia.

A partir de esta clasificación se puede calcular la sensibilidad o tasa de verdaderos positivos que se define como:

$$Sensibilidad = \frac{VP}{VP + FN}$$

y la especificidad o tasa de falsos positivos que se define como:

$$Especificidad = \frac{FP}{FP + VN}$$

Variando el valor del umbral se construye la curva ROC:

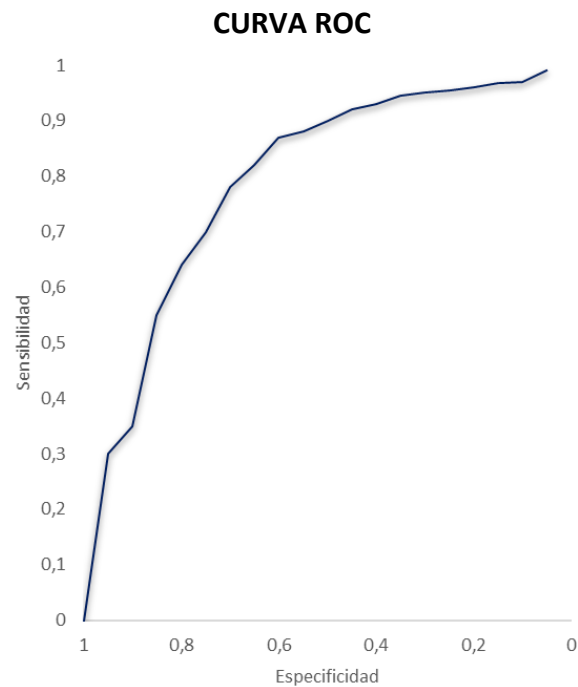


Gráfico 18. Fuente: elaboración propia.

Por último se calcula el área bajo la curva para obtener el AUC, buscando maximizar este valor por lo que al comparar algoritmos con esta métrica se selecciona aquel que tenga mayor valor.

## 6. Resultados

Teniendo en cuenta que el algoritmo seleccionado fue el método jerárquico se construyó el dendrograma correspondiente:

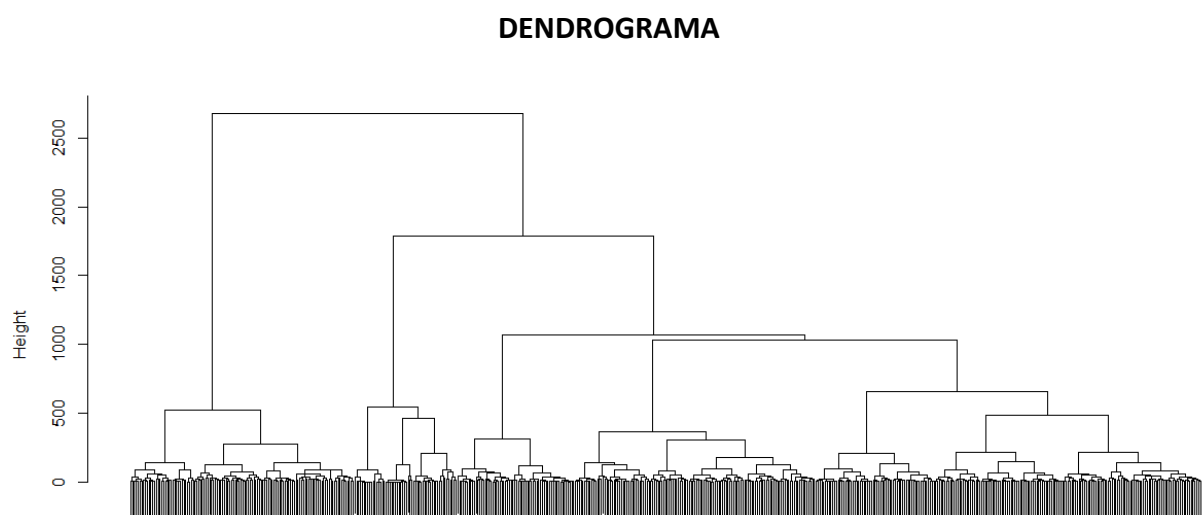


Gráfico 19. Fuente: elaboración propia.

Al observar las distancias entre los grupos, al igual que los resultados obtenidos al evaluar las medidas internas y de estabilidad se escogieron 3 segmentos para el análisis. Con esto se encontró que el primer segmento está conformado por 122 personas jurídicas y naturales, es decir, un 20.93% del total. Por otro lado, el segmento 3 está constituido por solo 9.61% que corresponde a 56 personas jurídicas y naturales, mientras que el segmento 2 lo integra el 69.47%, de forma que, con 405 de ellos este segmento es el más grande tal como se puede ver en el siguiente gráfico:

#### NÚMERO DE PERSONAS JURÍDICAS Y NATURALES POR SEGMENTO

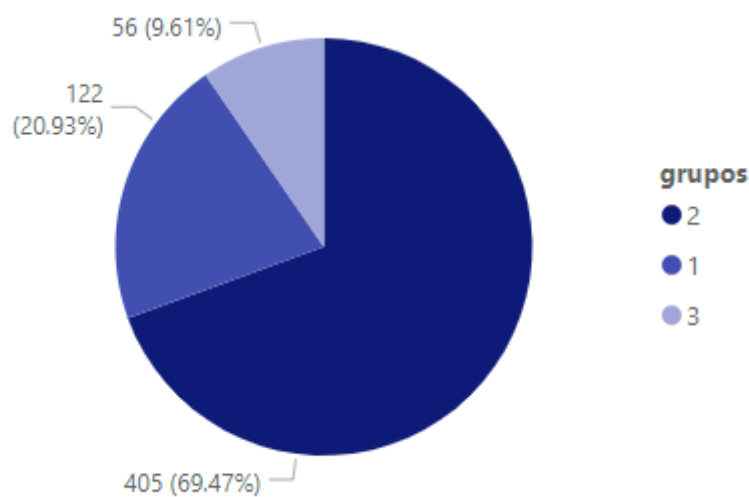


Gráfico 20. Fuente: elaboración propia.

Verificando las características que tienen cada uno de los segmentos se encontró que en el primer segmento se encuentran personas jurídicas y naturales que en promedio tienen 20 años de experiencia, más de 150 empleados, sus actividades económicas tienen una alta relación con construcción y adicionalmente allí se encuentran quienes ganaron los contratos que fueron analizados; en promedio ganaron cerca del 40% de los contratos a los que se presentaron. Por otro lado, en el segundo segmento se encuentran personas jurídicas y naturales que en promedio tienen aproximadamente 18 años de experiencia, cerca de 75 empleados, sus actividades económicas tienen una muy alta relación con construcción, pero no ganaron ninguno de los contratos que fueron analizados. En contraste, el tercer segmento está conformado por personas jurídicas y naturales que en promedio tienen alrededor de 10

años de experiencia, menos de 25 empleados, sus actividades económicas no se relacionan con construcción y no ganaron ninguno de los contratos a los que se presentaron.

### CARACTERIZACIÓN DE LOS SEGMENTOS

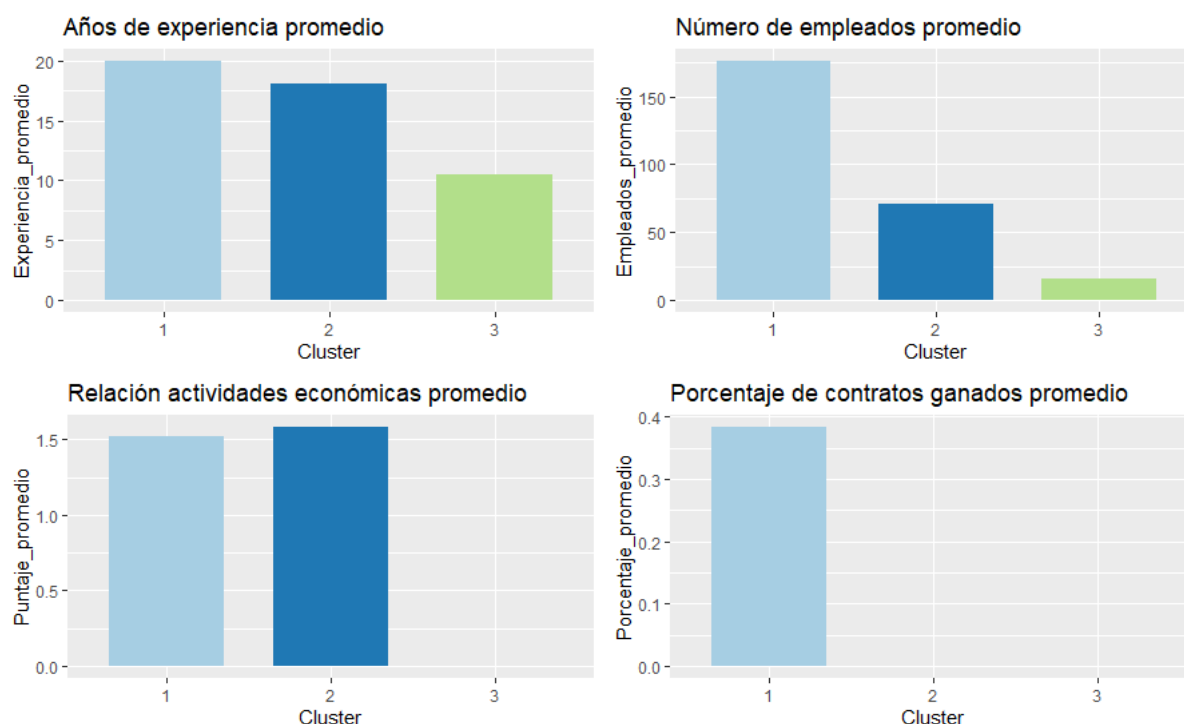


Gráfico 21. Fuente: elaboración propia.

Lo anterior muestra que en los dos primeros segmentos se encuentran las personas jurídicas y naturales con más experiencia y con una alta relación entre sus actividades económicas y la construcción, siendo ligeramente mayor para el segmento 2; sin embargo, al ver el número de empleados como una medida del tamaño de la empresa, muestra que el segmento 1 está conformado por personas jurídicas y naturales más grandes que el segmento 2. Considerando que se analizaron 65 contratos a los que se presentaron 583 personas jurídicas y naturales, es entendible que buena parte de ellos no ganaran ninguno de los contratos analizados; aun así, vale la pena observar qué ocurre al interior del segmento que concentra los ganadores. De la misma forma, es importante detallar el segmento 3 en donde se encuentran oferentes con menor experiencia, menos empleados y no se evidencia una relación entre sus actividades económicas y la construcción. Para este punto, se debe recordar que al calcular el puntaje asociado a la relación de las actividades económicas con construcción este podría

ser 0 si no tenía alguna relación o si no se encontró información relacionada con las actividades económicas.

En consecuencia, se revisaron los segmentos 1 y 3 con más detalle, para lo cual se analizó la forma en que se conformaban dichos segmentos. En el segmento 3 se encontraron personas jurídicas y naturales que resultaban particularmente interesantes, pues no se encontraron sus registros en las Cámaras de Comercio a pesar de que una búsqueda más amplia mostró que, al parecer, están ejecutando contratos con otras entidades públicas actualmente. Por otro lado, en el segmento 1 se pudo ver cómo se agrupan personas jurídicas y naturales directamente relacionadas con grandes grupos económicos. En dicho segmento se evidenció al menos un caso donde personas jurídicas y naturales pertenecientes a un mismo grupo económico se presentaban por separado a un mismo contrato, es decir, dicho grupo económico presentaba más de una propuesta para un contrato. Lo anterior puede indicar que los contratos se concentran en ciertos grupos de proponentes y que pueden existir prácticas que afectan las condiciones de competencia en estos procesos.

Finalmente, considerando las características de las personas jurídicas y naturales junto con los segmentos en los que fueron asignados, se calibraron dos algoritmos de clasificación (Random Forest y Boosting). Para esto la muestra se dividió de forma aleatoria entre los datos de entrenamiento y de prueba en una proporción de 60%-40% respectivamente. Al entrenar los algoritmos se encontró el siguiente resultado:

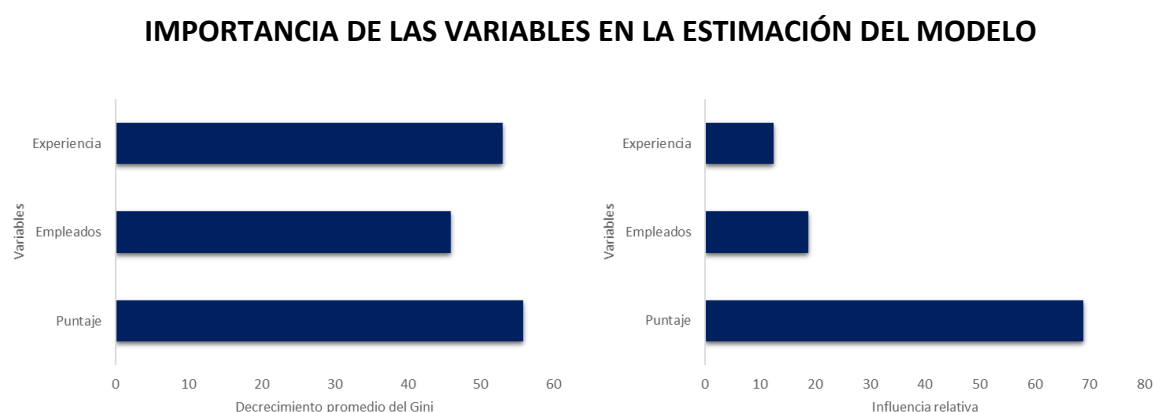


Gráfico 22. a) Random Forest | b) Boosting Fuente: elaboración propia

El puntaje asociado a la relación entre las actividades económicas y construcción en ambos modelos fue la variable que tuvo mayor importancia al entrenar los modelos, especialmente

en el caso de Boosting. Posteriormente, utilizando los datos de prueba se calculó el área bajo la curva ROC de los algoritmos y se encontró el siguiente resultado:

Algoritmo	AUC
Random Forest	0.84
Boosting	0.82

Por lo tanto, Random Forest es el algoritmo con mejor desempeño para este caso. El algoritmo permite estimar la probabilidad de que una nueva persona jurídica o natural se encuentre en uno de los tres segmentos a partir de sus años de experiencia, número de empleados y el puntaje de la relación entre sus actividades económicas y la construcción. De forma que es un signo de alerta: si el algoritmo indica que una persona jurídica o natural tiene mayor probabilidad de encontrarse en los segmentos 1 o 3 se recomienda revisar con más detalle los contratos en los que ha participado.

## 7. Conclusiones

Este trabajo tuvo como objetivo realizar un análisis de las relaciones entre los proponentes de contratos de vías primarias en Colombia, dirigido a facilitar la identificación de posibles casos que afecten la competitividad de dichos procesos a través de una herramienta de Machine Learning. Para esto se construyó una base datos con la información de los proponentes que participaron en 65 contratos de vías primarias que se ejecutaron entre los años 2014 y 2018 a nivel nacional. Adicionalmente, se caracterizó cada de los proponentes teniendo en cuenta los años de experiencia, el número de empleados y la relación entre las actividades económicas que registraban ante la Cámara de Comercio y la construcción de vías.

Posteriormente, se evaluaron 4 algoritmos de segmentación con un conjunto de métricas de dos clases (internas y de estabilidad) y se escogió el método jerárquico debido a su desempeño. Al implementar dicho algoritmo se encontró que la adjudicación de contratos de vías primarias se concentra en ciertos proponentes que están relacionados con los grandes grupos económicos del país. Esto último no refleja la pluralidad de proponentes que se esperaría dado que son contratos de grandes cuantías con un alto impacto en el desarrollo del país.

También se evidenció que dentro de los procesos pueden estar ocurriendo prácticas que afectan la competitividad en la contratación de vías primarias en el país. No solo porque la adjudicación de los contratos se concentra principalmente en los grandes grupos económicos, como se mencionó anteriormente, sino también porque se evidencian prácticas tal como que un mismo grupo económico presente más de una propuesta para un solo contrato. De forma que afecta la competitividad de los procesos pues si los grandes grupos económicos incurrían en este tipo de prácticas tienen ventajas frente a otros proponentes. Adicionalmente, se desarrolló un sistema de alertas utilizando un algoritmo de clasificación que estima la probabilidad de una persona jurídica o natural se encuentre en cada uno de los segmentos. Para este caso se entrenaron dos algoritmos de clasificación y utilizando el AUC como métrica de evaluación se seleccionó Random Forest puesto que presentaba mejor desempeño. Esto permite que para una nueva persona jurídica o natural, dada la información de años de experiencia, número de empleados y el puntaje de la relación de sus actividades económicas con construcción se pronostique el segmento al que pertenece. De esta forma se puede generar una alerta si es más probable que pertenezca a los segmentos 1 o 3 pues en estos segmentos se encontraban quienes podrían estar relacionados con prácticas que limitan la competencia en los procesos de contratación pública de vías primarias.

Por último, este trabajo permite ver cómo las herramientas de aprendizaje automático o Machine Learning pueden ayudar a las entidades de control a tener una mejor capacidad de respuesta. Para este caso, los algoritmos permiten identificar relaciones que pueden afectar la competitividad de los procesos de contratación y de esta forma reducir el número de proponentes y contratos que deben analizar.

## **8. Trabajo futuro**

Este trabajo puede continuar con la ampliación de los contratos de vías primarias que son analizados, pero adicionalmente vale la pena revisar los contratos relacionados o que se derivan de estos procesos, tales como los contratos de suministros e interventoría.

Este tipo de análisis también podría aplicarse a contratos con objetos distintos a los de construcción de vías para lo cual el SECOP II ofrece una gran oportunidad pues al ser un portal



transaccional facilita la trazabilidad de los procesos, reduciendo el esfuerzo que se requiere para organizar y limpiar la información disponible.

## 9. Bibliografía

Acdivoca.org.co. 2020. *Barómetro* – ACDI/VOCA. [online] Disponible en: <<http://www.acdivoca.org.co/barometro/>>

Bandiera , O., Prat, A., & Valletti, T. (2008). Active and Passive Waste in Government Spending: Evidence from a Policy Experiment. *American Economic Review*, 1278–1308.

Bohórquez, K. (23 de enero de 2019). Solo 13% de la red vial primaria está en muy buen estado. *La República*.

Breiman, Leo (2001). «Random Forests». *Machine Learning* 45 (1): 5-32. doi:10.1023/A:1010933404324.

Brock, G., Pihur, V., Datta, S. and Datta, S., 2008. clValid: AnRPackage for Cluster Validation. *Journal of Statistical Software*, 25(4).

Colombia Compra Eficiente. (2016). Guía para las Entidades Estatales con régimen especial de contratación.

Dunn, J. (1974). "Well separated clusters and optimal fuzzy partitions". *Journal of Cybernetics*. 4: 95–104. doi:10.1080/01969727408546059

E. Grace, A. Rai, E. Redmiles and R. Ghani, "Detecting fraud, corruption, and collusion in international development contracts: The design of a proof-of-concept automated system," *2016 IEEE International Conference on Big Data (Big Data)*, Washington, DC, 2016, pp. 1444-1453, doi: 10.1109/BigData.2016.7840752.

Galvis-Quintero, Deisy, La colusión como una práctica restrictiva de la competencia que afecta gravemente los procesos de selección de contratistas, 132 *Vniversitas*, 133-196 (2016). <http://dx.doi.org/10.11144/Javeriana.vj132.cprc>

Hanley, J. and McNeil, B., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), pp.29-36

Hastie, T., Tibshirani, R. and Friedman, J., 2009. *The Elements Of Statistical Learning*. Dordrecht: Springer.

INVÍAS. (29 de agosto de 2016). *Clasificación de las carreteras*. Obtenido de <http://www.invias.gov.co/index.php/2-uncategorised/2706-clasificacion-de-las-carreteras>

INVIAS. (s.f.). *Mapa de carreteras*. Obtenido de <https://hermes.invias.gov.co/carreteras/>

Issa, Hussein & Vasarhelyi, Miklos. (2011). Application of Anomaly Detection Techniques to Identify Fraudulent Refunds. *SSRN Electronic Journal*. 10.2139/ssrn.1910468.

Jaakko Hollmen (9 Marzo 1996). "Self-Organizing Map (SOM)". Aalto University.

MONITOR CIUDADANO DE LA CORRUPCIÓN. (2019). *ASÍ SE MUEVE LA CORRUPCIÓN*. Bogotá: Corporación Transparencia por Colombia.

López-Iturriaga, F.J., Sanz, I.P. Predicting Public Corruption with Neural Networks: An Analysis of Spanish Provinces. *Soc Indic Res* 140, 975–998 (2018). <https://doi.org/10.1007/s11205-017-1802-2>

OCDE. (2014). *Combatiendo la Colusión en los Procesos de Contratación Pública en Colombia*.

O. Irisova, "The cost of corruption," *World Economic Journal*, Agosto 2014.

Palacios Luna, J. M. (2014 ). Efectos de la corrupción sobre el crecimiento económico. Un análisis empírico internacional. *En Contexto*, 109 - 126.

SIC. (2013). *GUÍA PRÁCTICA Combatir la colusión en las licitaciones*.

Transparencia por Colombia. (23 de septiembre de 2019). *Barómetro Global de Corrupción para América Latina y el Caribe 2019*. Obtenido de <https://transparenciacolombia.org.co/2019/09/23/resultados-barometro-global-de-corrupcion-2019/>

Transparencia por Colombia. (02 de 05 de 2019). *Índice de Percepción de la Corrupción 2018*. Obtenido de <http://transparenciacolombia.org.co/2019/02/25/resultados-ipc-2018/>

Zuleta, M. M., Ospina, S., & Caro, C. A. (2019). *Índice de riesgo de corrupción en el sistema de compra pública colombiano a partir de una metodología desarrollada por el Instituto Mexicano para la Competitividad*. Banco Interamericano de Desarrollo.

Zuleta, M. M. & Caro, C. A, (2020). *Mirada a las alertas de riesgo de corrupción del sistema de compra pública colombiano identificadas en el cálculo del índice con la metodología del IMCO*. Bogotá: Fedesarrollo, 81 p. Laboratorio Latinoamericano de Políticas de Probidad y Transparencia. Un Proyecto de Cooperación Sur-Sur ATN O/C 16465-RG