

MODELO ANALITICO PARA ADACOP

MAESTRÍA EN ANALÍTICA PARA INTELIGENCIA DE NEGOCIOS



Profesores:

Luis Manuel Pulido Moreno

Alejandro Sierra Múnera

Presentado Por:

Oswaldo Martínez Marín

BOGOTÁ, 05 DE DICIEMBRE DE 2019

CONTENIDO

1. Introducción
2. Entendimiento del negocio
 - 2.1. Determinar Objetivos de negocios
 - 2.2. Evaluación de la situación
 - 2.2.1. Antecedentes
 - 2.2.2. Problemática
 - 2.3. Determinar objetivos de minería de datos
 - 2.3.1. Objetivo general
 - 2.3.2. Objetivos de minería de datos
 - 2.3.3. Éxito de la minería de datos
3. Entendimiento de los datos
 - 3.1. Recoger datos iniciales
 - 3.2. Describir datos
4. Informe
 - 4.1 Explorar datos
5. Preparación de los datos
 - 5.1. Seleccionar los datos
6. Modelamiento
 - 6.1. Escoger la técnica de modelamiento
7. Conclusiones

1. Introducción

Por invitación directa del Ministerio de Tecnologías de la Información y las Comunicaciones y el Departamento Administrativo de Ciencia, Tecnología e Innovación –Colciencias-, diferentes empresas del sector público y el sector privado de Colombia se unieron con el propósito de fortalecer la generación de soluciones de análisis de información.

Gracias a este acuerdo, nació el Centro de Excelencia y apropiación en Big Data y Data Analytics – Alianza CAOBA -, el cual tiene como objetivo generar soluciones en diversos sectores industriales, gubernamentales y académicos.

Este centro está constituido por las empresas Grupo Bancolombia, Grupo Nutresa, IBM de Colombia, SAS Institute Colombia, EMC Information Systems Colombia, Cluster CREATIC, Departamento Nacional de Planeación (DNP) y las Universidades ICESI, EAFIT, los Andes y la Pontificia Universidad Javeriana, que actúa como ejecutor del proyecto. Alianza CAOBA combina la capacidad de investigación de las universidades, desarrollando la oferta de productos líderes en tecnología a nivel mundial que tienen el objetivo de generar servicios y soluciones innovadores, que agreguen valor a los sectores estratégicos del país (www.alianzacaoba.co).

Este centro de excelencia apoya el uso de las tecnologías de Big Data y Data Analytics, a través de diferentes frentes que incluyen la formación de talento humano, la investigación aplicada y el desarrollo de productos, cuya propuesta de valor está fundamentada en la generación de soluciones alrededor de las tecnologías del BD&DA.

En el marco del Congreso Internacional TIC - ANDICOM 2018, la Ministra de Tecnologías de la Información y las Comunicaciones (TIC), Sylvia Constaín, presentó los ejes estructurales de la política TIC 2018 -2022: ‘El futuro digital es de todos’, con la cual se busca cerrar la brecha digital en Colombia. Esta política aborda las necesidades más apremiantes del país: conectar a todos los colombianos, solventando el déficit de redes de última milla; crear y fortalecer habilidades digitales en todos los ciudadanos, generar más negocios digitales y fortalecer el emprendimiento.

La ministra Constaín resaltó que esta tarea no puede ser asumida sólo por el Gobierno, sino que se necesita un trabajo conjunto con el sector privado y toda la ciudadanía. Por esto, una de las principales apuestas de esta política es crear una gobernanza acorde a las nuevas realidades, que dinamice la inversión privada, estrategia principal que busca beneficiar a todos los colombianos.

El MinTIC tiene definidos cuatro pilares, sobre los que se soporta la política “El futuro digital es de todos” (Mintic.gov.co. El Futuro digital es de todos), con los que se busca cerrar la brecha digital en Colombia. Los cuatro pilares son:

1. “Entorno TIC para el desarrollo digital”, con el que se busca modernizar el sector TIC;
2. “Ciudadanos y hogares empoderados del entorno digital”, cuyo objetivo es empoderar a las personas para que interactúen más de los servicios TIC;
3. “Inclusión social digital”, centrado en lograr una cobertura total de internet a lo largo del territorio nacional;
4. “Transformación digital sectorial y territorial”, para que los sectores público y privado exploten al máximo las tecnologías en sus procesos.

De acuerdo a lo anteriormente mencionado, CAOBA desarrollo la plataforma ADACOP que esta alineada a los pilares de MINTIC:

La plataforma ADACOP consiste en el análisis de Datos Abiertos relacionados con la Contratación Pública, es una herramienta desarrollada por el Centro de Excelencia y Apropiación en Big Data y Data Analytics - Alianza CAOBA que está disponible para que los ciudadanos y entidades públicas tengan un mejor entendimiento sobre las dinámicas relacionadas con el proceso de contratación estatal. Esta herramienta fue desarrollada de acuerdo a los pilares definidos por el Ministerio de Tecnologías de la Información y las Comunicaciones MinTIC, particularmente de acuerdo a los pilares de transparencia y empoderamiento al ciudadano. Todas las fuentes de datos de los componentes presentados provienen del Portal de Datos Abiertos.

ADACOP ofrece al usuario dos tipos de análisis. En primer lugar, se tiene un análisis sobre las dinámicas de los procesos de contratación en SECOP I. Con este componente es posible estudiar fenómenos como el flujo y origen de los recursos. además, de la temporalidad en términos de cuantías y cantidad de contratos. Adicionalmente, se ofrece un componente de análisis de calidad de datos con el propósito de tener un mejor entendimiento y disponer de una herramienta para la trazabilidad de las problemáticas existentes en la calidad de datos de la fuente de SECOP II.

La arquitectura de ADACOP fue diseñada para extender su uso al de otras fuentes que pertenecen al Portal de Datos Abiertos, por lo que próximamente se estará presentando al público una serie de análisis similares en otros campos de interés como la salud, la movilidad, la economía, entre otros.

Como se menciona anteriormente la plataforma ADACOP realiza un análisis descriptivo de la información, con el objetivo de robustecer la plataforma y poder ofrecer mayores herramientas para la toma de decisiones de los diferentes usuarios de esta. En este orden de ideas CAOBA junto al observatorio Jurídico de la Universidad Javeriana desea implementar herramientas analíticas dentro de la plataforma que permita establecer

cuando un contrato posiblemente llegara a sobrecostos y en que montos, adicionalmente, si presentara demoras en su ejecución, esto con el fin de empoderar a los usuarios (ciudadanos del común, entidades publicas y privadas) para que logren tomar mejor decisiones y que exista una mayor transparencia y entendimiento de la contratación estatal.

2. Entendimiento del negocio

2.1 Determinar Objetivos de negocios

Uno de los objetivos estratégicos de CAOBA es impulsar la gestión del conocimiento en analítica. Parte fundamental de la gestión del conocimiento es su efectiva transferencia, esta se ve fortalecido en el trabajo conjunto entre CAOBA y el Observatorio fiscal de la Pontificia Universidad Javeriana; identificando que herramientas analíticas (modelos) pueden apoyar la transparencia en la contratación pública del estado colombiano aportando a todos los actores involucrados (ciudadanos, entidades públicas, entre otros) un mayor entendimiento del proceso de contratación estatal, robusteciendo la actual plataforma ADACOP.

2.2 Evaluación de la situación

En los últimos años las principales economías del mundo han avanzado en el reconocimiento de los datos como elemento central para la adaptación hacia la cuarta revolución industrial y para el apalancamiento del crecimiento económico.

El panorama del uso y aprovechamiento de los datos por parte de las entidades publicas es desolador, como se puede observar tan solo el 9,3% de las entidades públicas tenían al menos un proyecto de aprovechamiento de datos en 2017, la meta al 2020 es que el 45% de ellas cuenten con al menos un proyecto.

Tabla 8. Indicadores de resultado de la política de explotación de datos

Indicadores de resultado	Línea base 2017	Meta 2020	Meta 2022
Porcentaje de entidades que tienen al menos un proyecto de aprovechamiento de datos	9,3 %	45 %	90 %
Entidades del orden nacional que avanzaron al tercer nivel del índice de preparación para la explotación de datos	8,7 %	20 %	50 %
Entidades del Gobierno que hacen proyectos de aprovechamiento de datos para mejorar servicios prestados a la ciudadanía	5,3 %	10 %	50 %
Promedio de activos públicos digitalizados y publicados	51 %	80 %	100 %

Fuente: DNP (2018).

Ahora, con el proyecto de Datos Abiertos, el Gobierno Colombiano promueve la transparencia, el acceso a la información pública, la competitividad, el desarrollo económico, y la generación de impacto social a través de la apertura, la reutilización de los datos públicos, y el uso y apropiación de las TIC.

La iniciativa de Datos Abiertos busca que todas las entidades del sector público publiquen la información pertinente y de calidad en formatos estructurados a disposición de los usuarios, para que ellos y las entidades la utilicen de diferentes maneras, según su interés: generar informes, reportes, estadísticas, investigaciones, control social, oportunidades de negocio (ej. aplicaciones), entre otros temas.

El concepto de Datos Abiertos es entendido como una práctica basada en la idea de que los datos o la información creados por la Administración Pública pertenezcan a la sociedad.

Este proyecto responde a las principales directrices de la construcción de Gobierno Abierto a nivel internacional. Colombia se adhirió a la Alianza para un Gobierno Abierto (AGA-OGP) en el año 2011. Allí se señaló como uno de los principales retos el: "Como integrantes de la Alianza para el Gobierno Abierto (...) Nos comprometemos a proporcionar activamente información de alto valor, incluidos los datos primarios, de manera oportuna, en formatos que el público pueda encontrar, comprender y utilizar fácilmente, y en formatos que faciliten su reutilización."

Por otra parte pero en la misma vía, El Departamento Nacional de Planeación y el Ministerio de Tecnologías de la Información y las Comunicaciones liderarán la implementación de la Política nacional de explotación de datos (Big Data), en un periodo de implementación de cinco años, con inversiones por 16.728 millones de pesos (Documento Conpes 3920) cuyo objetivo general consiste en aumentar el

aprovechamiento de datos en Colombia, mediante el desarrollo de las condiciones para que sean gestionados como activos y generar valor social y económico.

De acuerdo a lo anterior se puede evidenciar el esfuerzo del gobierno nacional para poner en marcha proyectos que permitirán que el país tenga un mayor nivel de madurez en la utilización de datos y en la toma de decisiones. En ese orden de ideas, este trabajo aportara respecto a lo mencioando anteriormente, la muestra de cómo sacarle provecho a los datos abiertos del SECOP sobre contratación pública genera valor y utilizacion eficaz de la información, debido que como se muestra a continuación aún se presentan fallas en el proceso.

2.2.1 Antecedentes

Durante el desarrollo del trabajo de grado de un grupo de estudiantes de la maestría en analítica de la universidad Javeriana, exploraron la posibilidad de tener modelos analíticos que permitieran perfilar los contratos con demoras y sobrecostos en la contratación pública, sin embargo, los resultados con las metodologías exploradas no fueron los esperados, aquí algunas de las conclusiones técnicas del estudio:

- De acuerdo a los dos modelos anteriores (Arboles de clasificación), en la predicción de demoras, teniendo en cuenta los resultados de la matriz de confusión y a pesar de que el valor del accuracy es superior al 88 % en ambos casos; los datos muestran que el porcentaje de instancias bien clasificadas (Precisión) es 0, lo que significa que en la predicción no hay contratos con demoras, a pesar de que la variable “objetivo” contaba con casos que justificaban que el contrato si estaba retrasado.
- En la predicción de demoras, los dos modelos generados presentan sobreajuste, dado que la clase dominante es 0, se pensó en solucionarlo realizando un balanceo de clases, pero la calidad de datos de SECOP en las variables seleccionadas no permitía generar un set de datos de entrenamiento y validación, nivelado.

El objetivo del presente trabajo es explorar de otra manera la construcción de modelos analíticos que permitan tener un mayor conocimiento sobre los tiempos y sobrecostos de los contratos del estado. Esto será posible al explorar distintas metodologías a las ya utilizadas y contando con información exógena adicional a la que ya se exploró.

2.2.2 Problemática

El Gobierno de Colombia ha reconocido que la compra y contratación pública es un asunto estratégico. Por este motivo, se creó la organización “Colombia Compra Eficiente” en el año 2011. Esta entidad crea políticas unificadas para que sirvan de guía a los administradores de compras del estado y que permitan monitorear y evaluar el desempeño del sistema generando mayor transparencia (Colombia Compra Eficiente, 2015).

Con la creación de esta organización, se adoptó el sistema de información SECOP (Sistema Electrónico para la Contratación Pública) promoviendo la contratación abierta y el uso de la información de los procesos de compra para fomentar la colaboración, la innovación y la transformación de la entrega de bienes, obras y servicios a los ciudadanos. Un proceso de contratación comprende una serie de pasos para que las entidades estatales puedan adquirir los bienes y servicios, mediante las licitaciones en las que participan los proveedores. Los pasos se resumen de la siguiente manera:

1. Una entidad estatal abre un proceso de contratación para adquirir bienes o servicio.
2. Los proveedores que deseen contratar con el estado se registran en la plataforma y son clasificados de acuerdo con el tipo o bien que ofrecen.
3. Una vez se publica el proceso de contratación, los proveedores reciben una invitación de acuerdo con el servicio/bien que ofrecen.
4. Tras la publicación del proceso, los proveedores manifiestan el interés por participar.
5. Finalmente, los proveedores que presentaron sus ofertas son tenidos en cuenta por la entidad quien selecciona uno de ellos.



*Con base en información de 2016 en adelante

La **contratación pública** es realizada por **Entidades del Estado** que realizan compras de **Bienes y Servicios** necesarias para su correcto funcionamiento.

Los detalles de los contratos **deben ser registrados** en el sistema de compra pública de **Colombia Compra Eficiente**



Los **detalles de las compras** deben ser registrados en el **sistema de contratación pública** dispuesto por **Colombia Compra Eficiente (CCE)** a través de sus tres plataformas:

- Secop I
- Secop II
- Tienda Virtual del estado Colombiano

El gobierno nacional ha dado pasos importantes en la utilización de tecnologías de la información para poder gestionar de manera efectiva los datos públicos desde la elaboración de políticas públicas como el conpes de BIG DATA, y el proyecto de datos abiertos, pero aún existen muchas brechas que hay que comenzar a cerrar sobre todo lo concerniente a la utilización y exploración de los datos con el fin de poder tener una mejor gestión y entendimiento del Estado y lograr así pasar de solo gestión efectiva de la información a conocimiento explícito sobre la sociedad en general.

Evaluar la situación

Inventario de recursos

Personal

1 estudiantes de maestría en Analítica de inteligencia de negocios

1 conocedores del negocio

1 intermediario universidad – entidad

2 profesores

Datos

Base de datos abiertos del SECOP sobre contratación pública.

Máquina, plataforma y software

Programa R, KNIME

Equipos portátiles

Requisitos, Supuestos, y Restricciones

Requisitos:

Para la elaboración de este modelo analítico se hace necesario:

- Contar con el equipo de cómputo y software requerido para el procesamiento de los datos
 - Personas con el conocimiento necesario para realizar los modelos requeridos (Estudiante)

Supuestos:

- La información puede ser accedida en su totalidad ya que son datos abiertos.
- No se requiere muchas tareas de limpieza para tener una base de datos lista para la generación de los modelos

Restricciones:

- Acceso a ciertas variables que no se encuentren en datos abiertos del SECOP.
- Restricciones de tiempo debido a las demoras ocurridas en el inicio de este proyecto (1 Mes y medio) desde el inicio del semestre académico.

Riesgos y Contingencias

A nivel de volumen de datos y calidad, el riesgo es que la información proporcionada no sea lo suficientemente representativa para poder generar un modelo con la precisión que se espera.

Terminología

Terminología del negocio

Datos abiertos: son “datos primarios o sin procesar, que se encuentran en formatos estándar e interoperables que facilitan su acceso y reutilización, los cuales están bajo la custodia de las entidades públicas o privadas que cumplen con funciones públicas y que son puestos a disposición de cualquier ciudadano, de forma libre y sin restricciones, con el fin de que terceros puedan reutilizarlos y crear servicios derivados de los mismos”¹³⁹. La definición de apertura involucra los siguientes elementos: (i) disponibilidad y acceso¹⁴⁰; (ii) ¹³⁸ Por medio de la cual se dicta la Ley General de Archivos y se dictan otras disposiciones. ¹³⁹ Literal j) del artículo 6 de la Ley 1712 de 2014. ¹⁴⁰ La información debe estar disponible como un todo y a un costo razonable de reproducción, preferiblemente descargándola de internet. Además, la información debe estar disponible en una forma conveniente y modificable. ⁹⁸ reutilización y redistribución¹⁴¹; y (iii) participación universal ¹⁴² (OKFN, 2016). La definición de apertura implica la exclusión de todos aquellos datos que, dada su tipología, presentan limitaciones jurídicas para su acceso, divulgación, compartición y tratamiento.

Gobierno Abierto: componente del Gobierno Electrónico que busca la construcción de un Estado más transparente, participativo y colaborativo en los asuntos públicos mediante el uso de las TIC.

2.3 Determinar objetivos de minería de datos

2.3.1 Objetivo general

Poder contar con modelos analíticos que permita explorar y explotar de manera eficiente los datos abiertos de la contratación pública en Colombia, esto con el fin de poder ofrecerles a la sociedad mayor entendimiento de la contratación pública con herramientas analíticas propias de la revolución 4.0.

2.3.2 Objetivos de minería de datos

- Explorar la información de forma descriptiva para poder evaluar la calidad de la información.
- Realizar análisis descriptivo para tener un entendimiento general del comportamiento de los contratos en Colombia.
- Construir modelos analíticos que permita determinar las características que tienen los contratos para caer en demoras y sobrecostos.

2.3.3 Éxito en la minería de datos

Poder contar con modelos analíticos que tengan métricas de ajustes adecuadas y que permitan explorar que características tienen un impacto significativo en la demora y sobre costo de los contratos.

3. Entendimiento de los datos

3.1 Recoger datos iniciales

Recopilación inicial de datos

Se cuenta con disponibilidad inmediata de la información en el portal de datos establecido por el gobierno, el cual se encuentra el link a continuación: <https://www.datos.gov.co/Gastos-Gubernamentales/SECOP-II-Contratos/gnxj-bape>

Conjunto de Datos

SECOP II-Contratos

Información de los Contratos correspondientes a procesos de compra pública registrados en la plataforma SECOP II desde su existencia.

Los metadatos de este conjunto de datos se encuentran en Anexo I, este conjunto de datos contiene 354K filas y 40 columnas, cada fila representa un contrato.

SECOP I

Información de los procesos de compra pública registrados en la plataforma SECOP I a partir del año 2011.

Los metadatos de este conjunto de datos se encuentran en Anexo II, este conjunto de datos contiene 7.8 millones filas y 62 columnas.

3.2 Describir datos

Descripción de los datos

En SECOP II-CONTRATOS las fechas de interés para poder determinar si el contrato tuvo o no demora son:

Nombre Campo	Tipo de dato	Descripción
Fecha de Inicio del Contrato	Fecha	Fecha de Inicio del contrato según el registro de la entidad, es la fecha base para cambios de estado
Fecha de Fin del Contrato	Fecha	Fecha de fin del contrato según el registro de la entidad en la plataforma, es la fecha base para cambios de estado como la liquidación y cierre
Fecha de Inicio de Ejecución	Fecha	Fecha de Inicio de la ejecución de las actividades o de entrega de los bienes definidos en el contrato
Fecha de Fin de Ejecución	Fecha	Fecha de fin de la ejecución de las actividades o de entrega de los bienes definidos en el contrato

A partir de esta información se construirá un nuevo atributo llamado:

TIEMPO_DEMORA= Fecha de Fin de Ejecución-Fecha de Fin del Contrato

Este tiempo estará expresado en días.

En SECOP I se utilizarán las siguientes métricas que me determina si el proceso esta en demora:

Tiempo Adiciones en Días	Entero	Extensión del contrato, fuera de la definición inicial, en días
Tiempo Adiciones en Meses	Entero	Extensión del contrato, fuera de la definición inicial, en meses

Algunas variables de interés y de carácter binario son:

Nombre Campo	Tipo de dato	Descripción
Es Grupo	Texto	Marcador que determina si el proveedor adjudicado en el contrato es un grupo o unión temporal
Estado Contrato	Texto	Estado del contrato (Ej. Firmado, Modificado, Cerrado, Liquidado)
Habilita Pago Adelantado	Texto	Marcador que indica si el Contrato permite hacer pagos adelantados
Liquidación	Texto	Marcador que indica si el Contrato puede liquidarse
Obligaciones Ambientales	Texto	Marcador que indica si el Contrato tiene algún tipo de obligación de carácter ambiental
Obligaciones Posconsumo	Texto	Marcador que indica si el Contrato tiene algún tipo de obligación posterior al uso del bien o del consumo del producto

En el Anexo 1 se encuentra el diccionario de datos completo del SECOP, donde se describe una a una las variables de la base.

4. Informe

4.1 Explorar datos

Exploración y calidad de datos

La base de datos del SECOP II cuenta con información de 353.503 contratos, se exploró la calidad de la información y se definieron algunas reglas para poder identificar posibles inconsistencias. Las reglas establecidas son:

Regla 1: MISSING \$Fecha de Inicio de Ejecucion\$ =>TRUE

Se excluyen los contratos en los cuales no existe información de inicio de ejecución, una vez aplicado este filtro quedan 243.962 contratos

Regla 2: $\$Fecha\ de\ Fin\ de\ Ejecucion\ \leq\ \$Fecha\ de\ Inicio\ del\ Contrato\ \$\ OR\ \$Fecha\ de\ Fin\ de\ Ejecucion\ \leq\ \$Fecha\ de\ Inicio\ de\ Ejecucion\ \$\Rightarrow TRUE$

La regla 2 identifica datos inconsistentes ya que la fecha de fin de ejecución no puede ser inferior a la del inicio del contrato o la fecha fin de ejecución estar antes de la fecha de inicio de ejecución. Al aplicar este filtro la base se reduce a 241.348 contratos.

Regla 3: $MIS\ SING\ \$Fecha\ de\ Fin\ de\ Ejecucion\ \$\ OR\ \$Fecha\ de\ Fin\ del\ Contrato\ \$\geq\ "2019-10-08"\ OR\ \$Fecha\ de\ Fin\ de\ Ejecucion\ \$\geq\ "2019-10-08"\Rightarrow 1$

Donde la fecha 2019-10-08 es la fecha de descarga de la información.

Esta regla identifica 69.744 contratos de los cuales no se cuenta información de finalización del contrato, una vez retirado los contratos censurados nos quedamos con 171.760.

A continuación, se resume el proceso de depuración de la base de datos.

Regla	contratos	% contratos con respecto a la base inicial
0	353.503	100%
1	243.962	69%
2	241.348	68%
3	171.760	49%

Tabla 1. Depuración de la base de datos con las reglas de negocio diseñadas

La mayoría de los datos con una posible censura corresponden a contratos que su fecha de fin de ejecución es la fecha que representa un día en el futuro, es decir no ha sucedido así que se asume como una falla en la calidad de la información.

Estado del Contrato	# contratos	%
Activo	85.903	50%
Modificado	46.570	27%
terminado	10.390	6%
Cerrado	9.498	6%
Borrador	5.533	3%
enviado Proveedor	5.522	3%
En aprobación	4.507	3%
Cancelado	1.945	1%
cedido	1.500	1%
Suspendido	292	0%
Prorrogado	99	0%
Total	171.759	100%

Ilustración 1. Distribución de contratos por su Estado

De acuerdo a lo consultado con CAOBA todos los estados del contrato deberían incluirse en el estudio, sin embargo, se trató de indagar con el observatorio jurídico de la Universidad y no se tuvo respuesta alguna. Por tal razón se decide no tener en cuenta los siguientes tipos de contratos: Cancelados y Suspendidos.

Tipo de contrato	# Contratos	%
Prestación de servicios	137.919	80,30%
Compraventa	11.849	6,90%
Suministros	9.721	5,66%
Otro	7.504	4,37%
Obra	1.656	0,96%
Arrendamiento de inmuebles	1.276	0,74%
Consultoria	574	0,33%
Interventoria	539	0,31%
Asociación Público Privada	232	0,14%
Seguros	211	0,12%
No Especificado	139	0,08%
Arrendamiento de muebles	67	0,04%
Comision	39	0,02%
Venta muebles	16	0,01%
Acuerdo de cooperación	6	0,00%
Negocio fiduciario	6	0,00%
Servicios financieros	5	0,00%
Concesión	1	0,00%
Total	171.760	100,00%

Ilustración 2. Distribución de contratos por tipo

El 80,30% de los contratos corresponden a prestación de servicios, seguido de un casi 7 % de contratos asociados a compraventa.

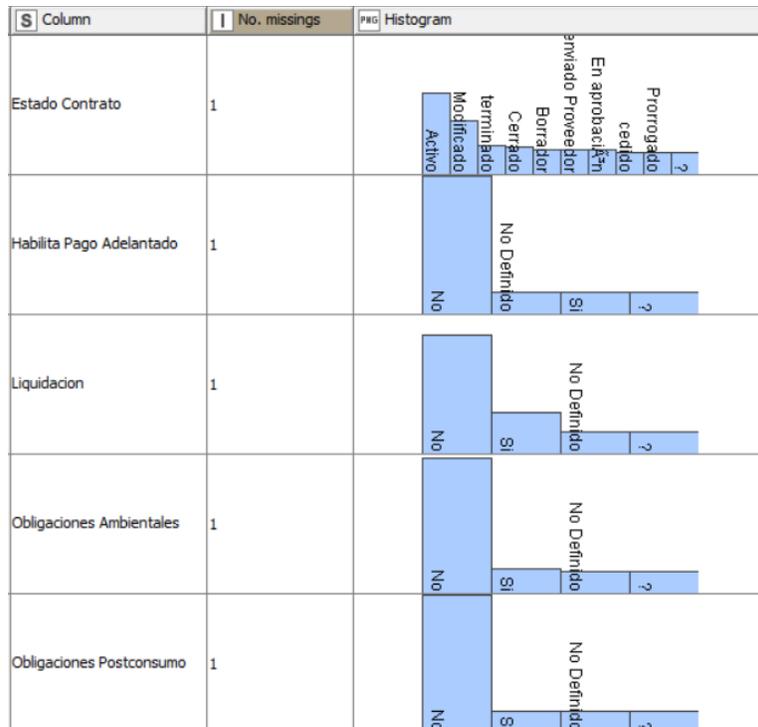


Ilustración 3. Completitud base de datos y distribución.

Como se puede observar en la ilustración 3, las variables que sería útiles en la construcción del modelo y después de la depuración de los datos cuentan con una calidad buena en términos de completitud (Ausencia de datos perdidos).

5. Preparación de los datos

Seleccionar los datos

Los datos que serán usados para cumplir con los objetivos de la minería de datos y del negocio corresponde a:

Variable objetivo: PRESENTA DEMORA (SI o NO)

Variables exógenas presentes en la base del SECOP:

- Es Grupo, Habilita Pago Adelantado, Liquidación, Obligaciones Ambientales y Obligaciones Posconsumo son variables con valores de SI o NO, estas son configuradas como variables binarias (1 y 0)
- Estado Contrato y tipo de contrato, estas variables categorías se tendrán en cuenta para el modelo.

Integración de datos

Debido a la poca información en las bases del SECOP que puedan dar explicación de cuando un contrato cae o no en demoras, se exploró algunas fuentes externas de información con el fin de poder contar con predictores más eficientes, las fuentes consultadas fueron:

Índice de transparencia de las entidades públicas (ITEP)

El Índice de Transparencia de las Entidades Públicas (ITEP) corresponde a una iniciativa que busca poder contribuir a la disminución de casos de corrupción en entidades del gobierno nacional.

Se define por lo tanto la corrupción como “el abuso de poder o de confianza por parte de un actor para obtener beneficios personales o de un grupo determinado de poder, en detrimento de los intereses colectivos”. Mientras que los riesgos de corrupción administrativa hacen referencia a “condiciones institucionales y prácticas de las autoridades que pueden favorecer la ocurrencia de hechos de corrupción”.

Además, la realidad indica que los riesgos de corrupción aumentan mientras el desarrollo de las entidades en temas de normas y protocolos son bajos.

Qué evalúa el ITEP

El índice evalúa fundamentalmente tres características vitales para controlar la corrupción en la administración pública:

Visibilidad: Corresponde a la competencia de una organización para hacer públicas sus políticas, procedimientos y decisiones, de manera suficiente, oportuna, clara y adecuada.

Institucionalidad: Corresponde a la competencia de una entidad para lograr que los servidores públicos y la administración en su conjunto cumplan con normas y estándares establecidos para los procesos de gestión.

Control y sanción: Corresponde a la competencia para generar acciones de control y sanción mediante procesos internos, por acción de los órganos de control y espacios de participación ciudadana.

Cada uno de los anteriores factores lo compone un conjunto de variables que centran su indagación en procesos claves para la gestión institucional.

Cómo se obtiene la calificación del indicador

Cada uno de los indicadores tienen pesos los cuales son asignados de acuerdo a los siguientes criterios:

- La importancia de la variable medida en temas de transparencia en procesos de gestión pública.
- La actualización de todas las herramientas, instrumentos tanto que existan como en su contenido y que aporten a la transparencia.

De acuerdo con lo anterior, “los factores de Visibilidad y Control y Sanción tienen un peso de 30 por ciento cada uno sobre la calificación final del Índice, mientras que el factor de Institucionalidad cuenta con un peso de 40 por ciento. Por su parte, los indicadores de cada factor también cuentan con una ponderación diferenciada”.

El Índice, así como cada unidad de medición cuenta con una calificación específica que va de cero (0) a cien (100) siendo cien (100) la mayor calificación posible.

El índice general de transparencia será identificado como ITB.

Esta fuente exógena de información presenta algunos inconvenientes como:

* La integración de la base de datos que contiene el índice de transparencia no es inmediata debido a que no existe una llave numérica como el NIT que esté presente en ambas bases. Por esto se realizó un match con métricas de similitud que permiten determinar cuál es el texto, en este caso, el nombre de la entidad que más son similares en las dos bases y realizar el match.

Se realiza una validación manual para verificar que se realizara adecuadamente el match entre las dos bases, además de verificar que las distancias calculadas mediante la métrica de Jaro-Winkler no fuera superior a

$$\left[\frac{\max(|s_1|, |s_2|)}{2} \right] - 1$$

Donde $|s_1|$ y $|s_2|$ corresponde a la longitud del nombre de la entidad en cada una de las fuentes de información.

S	Entidad	I	nearest neighbor - index	S	nearest neighbor - i→Nombre Entidad	D	distance
	Agencia Presidencial para la Cooperación Internacional	...	0		AGENCIA PRESIDENCIAL DE COOPERACION INT...		0.073
	Ministerio de Justicia y del derecho	...	0		MINISTERIO DE JUSTICIA Y DEL DERECHO		0
	Ministerio de Justicia y del Derecho	...	0		MINISTERIO DE JUSTICIA Y DEL DERECHO		0
	Ministerio de Hacienda y Crédito Público	...	0		MINISTERIO DE HACIENDA Y CREDITO PUBLICO		0.013
	Ministerio de Hacienda y Crédito Público	...	0		MINISTERIO DE HACIENDA Y CREDITO PUBLICO		0.013
	Ministerio de Educación Nacional	...	0		MINISTERIO DE EDUCACION NACIONAL (MEN)		0.041
	Ministerio de Educación Nacional	...	0		MINISTERIO DE EDUCACION NACIONAL (MEN)		0.041
	Ministerio de Defensa Nacional	...	0		MINISTERIO DE DEFENSA NACIONAL		0
	Ministerio de Defensa Nacional	...	0		MINISTERIO DE DEFENSA NACIONAL		0
	Ministerio de Cultura	...	0		MINISTERIO DE CULTURA		0
	Ministerio de Cultura	...	0		MINISTERIO DE CULTURA		0
	Agencia Presidencial para la Cooperación Internacional	...	0		AGENCIA PRESIDENCIAL DE COOPERACION INT...		0.073
	Ministerio de Comercio, Industria y Turismo	...	0		MINISTERIO DE COMERCIO INDUSTRIA Y TURIS...		0.003
	Ministerio de Ambiente y Desarrollo Sostenible	...	0		MINISTERIO DE AMBIENTE Y DESARROLLO SOS...		0
	Ministerio de Agricultura y Desarrollo Rural	...	0		MINISTERIO DE AGRICULTURA Y DESARROLLO ...		0
	Ministerio de Agricultura y Desarrollo Rural	...	0		MINISTERIO DE AGRICULTURA Y DESARROLLO ...		0
	Jardín Botánico José Celestino Mutis	...	0		JARDÍN BOTÁNICO JOSÉ CELESTINO M...		0.038
	Instituto para la Protección de la Niñez y la Juventud - IDIPRON	...	0		INSTITUTO DISTRITAL PARA LA PROTECCIÓN ...		0.08
	Instituto para la Investigación Educativa y el Desarrollo Pedagógico -	0		INSTITUTO PARA LA INVESTIGACIÓN EDUCATI...		0.01
	Instituto para la Economía Social - IPES	...	0		INSTITUTO PARA LA ECONOMIA SOCIAL- IPES		0.026
	Instituto de Planificación y Promoción de Soluciones Energéticas para	0		INSTITUTO DE PLANIFICACIÓN Y PROMOCIÓN...		0.014
	Instituto de Hidrología, Meteorología y Estudios Ambientales - IDEAM	...	0		INSTITUTO DE HIDROLOGÍA METEOROLÓGIA...		0.07
	Agencia Nacional de Minería	...	0		AGENCIA NACIONAL DE MINERIA		0.009
	Instituto de Desarrollo Urbano - IDU	...	0		INSTITUTO DE DESARROLLO URBANO		0.021
	Instituto Nacional de Vías - INVÍAS	...	0		INSTITUTO NACIONAL DE VIAS		0.04
	Instituto Nacional de Vías - INVÍAS	...	0		INSTITUTO NACIONAL DE VIAS		0.04
	Instituto Nacional de Vigilancia de Medicamentos y Alimentos - INVIMA	...	0		INSTITUTO NACIONAL DE VIGILANCIA DE MEDI...		0.017
	Instituto Nacional de Salud - INS	...	0		INSTITUTO NACIONAL DE SALUD		0.023
	Instituto Nacional de Medicina Legal y Ciencias Forenses	...	0		INSTITUTO NACIONAL DE MEDICINA LEGAL Y CI...		0
	Instituto Nacional Penitenciario y Carcelario - INPEC	...	0		INSTITUTO NACIONAL PENITENCIARIO Y CARC...		0.005
	Instituto Nacional De Salud - INS	...	0		INSTITUTO NACIONAL DE SALUD		0.023
	Instituto Geográfico Agustín Codazzi - IGAC	...	0		INSTITUTO GEOGRÁFICO AGUSTÍN CODAZZI		0.04
	Instituto Distrital de las Artes - IDARTES	...	0		INSTITUTO DISTRITAL DE LAS ARTES		0.03
	Agencia Nacional de Infraestructura - ANI	...	0		AGENCIA NACIONAL DE INFRAESTRUCTURA - ANI		0
	Instituto Distrital de la Participación y Acción Comunal - IDPAC	...	0		INSTITUTO DISTRITAL DE LA PARTICIPACIÓN Y...		0.029
	Instituto Distrital de Turismo - IDT	...	0		INSTITUTO DISTRITAL DE TURISMO		0.021
	Instituto Distrital de Recreación y Deporte - IDRD	...	0		INSTITUTO DISTRITAL DE RECREACIÓN Y DEP...		0.026
	Instituto Distrital de Patrimonio Cultural - IDPC	...	0		INSTITUTO DISTRITAL DE PATRIMONIO CULTURAL		0.021

Ilustración 4. Homologación del Nombre de la entidad SECOP Índice de transparencia usando métricas de similitud

No todas las entidades que aparecen en SECOP tienen calculado el índice de transparencia, en su mayoría son entidades del orden nacional que se encuentran en la capital colombiana, se cuenta con información de 167 entidades que tienen información del índice anual hasta del 2015 al 2017.

Debido a que no se cuenta con más información exógena que pueda ser obtenida en el transcurso del desarrollo de este estudio se decide tomar como base de estudio estas 167 entidades de orden distrital y nacional y se tomaran todos los contratos que se encuentran registrados en el SECOP. Estas 167 entidades representan 163.193 de los 171.760 es decir un 95% del total de contratos, la llave para el “join” corresponde a la entidad y el año de inicio de ejecución del contrato para el SECOP y para el índice de transparencia la entidad y el año de evaluación. Cabe aclarar que para los años 2018 y 2019 no existe información del ITB por lo tanto se asume para estos dos años el ultimo registrado.

A continuación, se realizan algunos gráficos descriptivos para observar la distribución del ITB y de los principales indicadores que los componen.

La grafica roja representa los contratos con demora y la azul sin demora. El primer grafico representa la distribución del ITB, aquí se observa que las entidades que tienen un puntaje mayor a 62 comienzan a tener una mayor probabilidad de no contar con demoras en los contratos. Esto sugiere la elaboración de una variable dummy:

ITB>62 ENTONCES ITB_DUMMY=1

NOT ITB>62 ENTONCES ITB_DUMMY=0

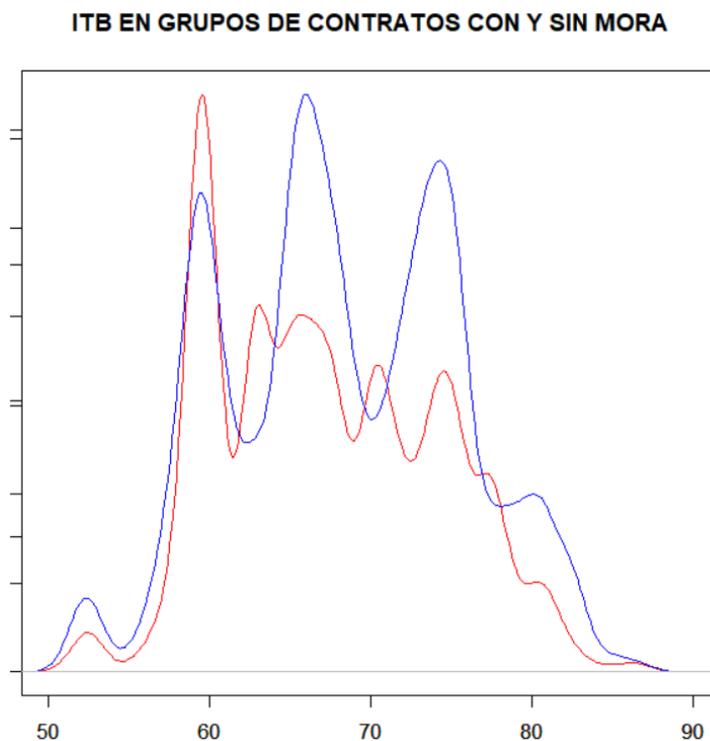


Ilustración 5. Distribución probabilística del IBT para contratos en demora (ROJO) y terminados a tiempo (AZUL)

La distribución del control y sanción es muy similar en los contratos con y sin mora, existe, una diferencia entre los puntajes menores a 55 ya que es más probable que el contrato caiga en demoras.

Esto sugiere la elaboración de una variable dummy:

CONTROL Y SANCIÓN<=55 ENTONCES CYS_DUMMY=1

NOT CONTROL Y SANCIÓN<=55 ENTONCES CYS_DUMMY=0

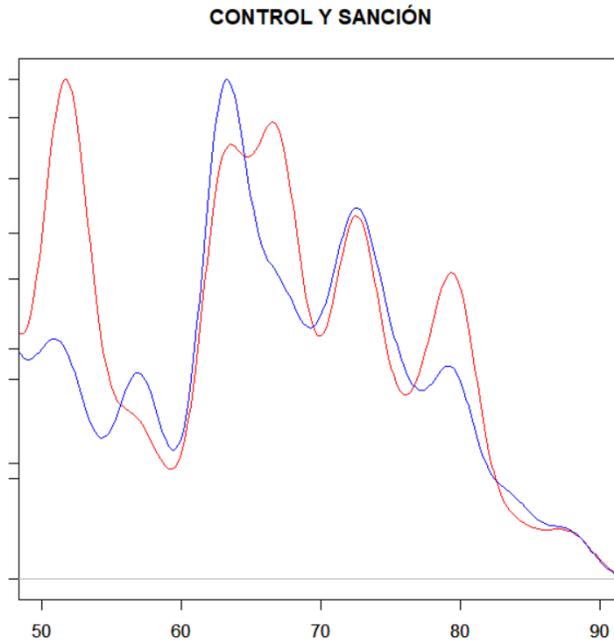


Ilustración 6. Distribución probabilística del indicador de Control y Sanción para contratos en demora (ROJO) y terminados a tiempo (AZUL)

En temas de institucionalidad la distribución del indicador es muy similar para los contratos que caen en demoras y los que no.

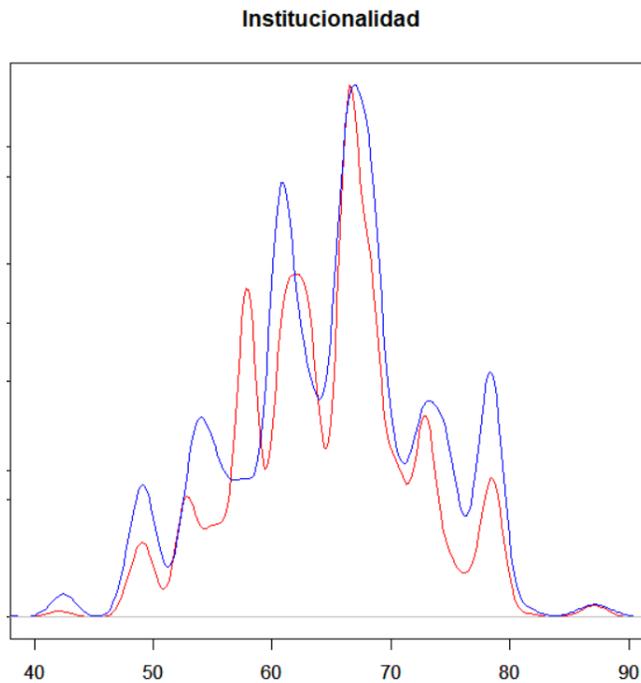


Ilustración 7. Distribución probabilística del indicador de Institucionalidad para contratos en demora (ROJO) y terminados a tiempo (AZUL)

En temas de condiciones institucionales de la información pública se puede evidenciar que para puntajes menores a 58 la probabilidad que caiga el contrato en mora es mayor a que no caiga en mora.

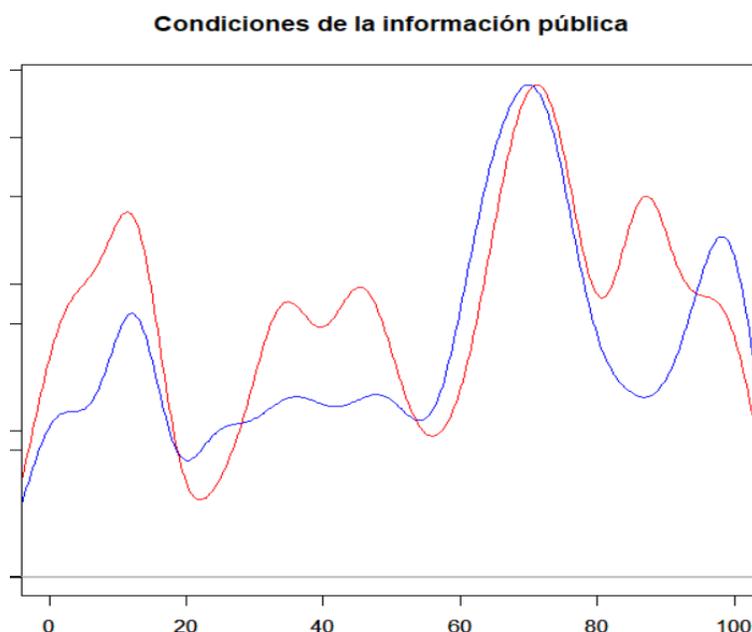


Ilustración 8. Distribución probabilística del indicador de condiciones institucionales de la información pública para contratos en demora (ROJO) y terminados a tiempo (AZUL)

Esto sugiere la elaboración de una variable dummy:

INFORMACION PUBLICA \leq 58 ENTONCES DUMMY_INFOPUBLICA=1

NOT INFORMACION PUBLICA \leq 58 ENTONCES DUMMY_INFOPUBLICA =0

En el indicador de visibilidad se puede concluir que si la entidad tiene un puntaje mayor a 72 la probabilidad de caer en demora es menor al de no caer.

Esto sugiere la elaboración de una variable dummy:

VISIBILIDAD \geq 72 ENTONCES VIS_DUMMY=1

NOT VISIBILIDAD \geq 72 ENTONCES VIS_DUMMY=0

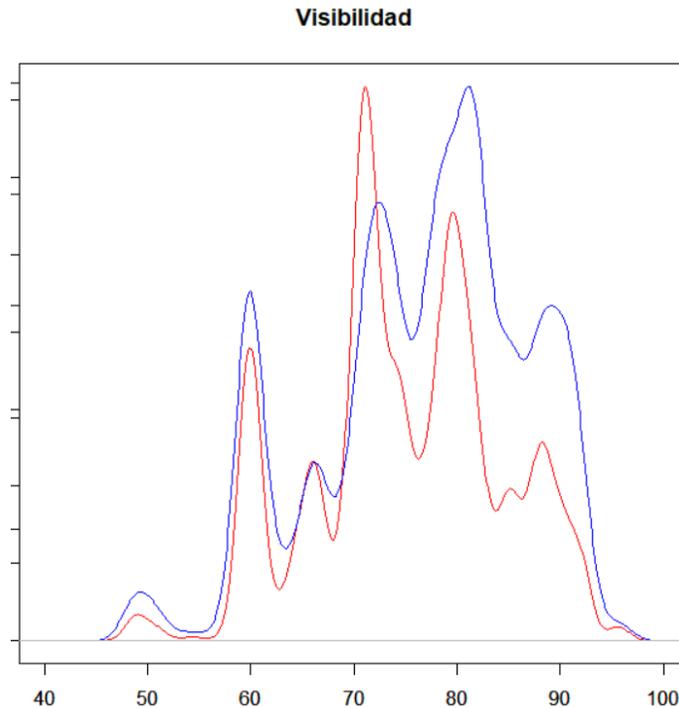


Ilustración 9. Distribución probabilística del indicador de Visibilidad para contratos en demora (ROJO) y terminados a tiempo (AZUL)

En la base actual se cuenta con información de 163.193 contratos de los cuales 147.678 corresponden a contratos que no caen en demoras y 15.515 que sí, debido a que las dos clases se encuentran desbalanceadas y con fines de evitar problemas como sesgos se seleccionan mediante un muestreo aleatorio simple de la clase de contratos que no caen en mora 15,515. Dado este balance se realiza la siguiente tabla de contingencia con las variables binarias descritas en la sección 2.2.

4. Modelamiento

5.1 Escoger la técnica de modelamiento

Debido a que el enfoque que tomo este proyecto es investigativo, se buscara la posibilidad de ajustar un modelo estadístico que permita dar **explicación** al comportamiento de interés, por tal razón se decide utilizar una regresión logística para modelar la caída en demora de un contrato.

La regresión logística (RL) forma parte del conjunto de métodos estadísticos que caen bajo tal denominación y es la variante que corresponde al caso en que se valora la

contribución de diferentes factores en la ocurrencia de un evento simple. En general, la regresión logística es adecuada cuando la variable de respuesta Y es polinómica (admite varias categorías de respuesta, tales como mejora mucho, empeora, se mantiene, mejora, mejora mucho), pero es especialmente útil en particular cuando solo hay dos posibles respuestas (cuando la variable de respuesta es dicotómica), que es el caso más común. La RL es una de las técnicas estadístico-inferenciales más empleadas en la producción científica contemporánea. Surge en la década del 60, su generalización dependía de la solución que se diera al problema de la estimación de los coeficientes. El algoritmo de Walker-Duncan para la obtención de los estimadores de máxima verosimilitud vino a solucionar en parte este problema, pero era de naturaleza tal que el uso de computadoras era imprescindible. La RL va a contestar a preguntas tales como: ¿Se puede predecir con antelación si un cliente que solicita un préstamo a un banco va a ser un cliente moroso? ¿Se puede predecir si una empresa va a entrar en bancarrota? ¿Se puede predecir de antemano que un paciente corra riesgo de un infarto? La identificación del mejor modelo de regresión logística se realiza mediante la comparación de modelos utilizando el cociente de verosimilitud, que indica a partir de los datos de la muestra cuanto más probable es un modelo frente al otro. La diferencia de los cocientes de verosimilitud entre dos modelos se distribuye según la ley de la Chi-cuadrado con los grados de libertad correspondientes a la diferencia en el número de variables entre ambos modelos. Si a partir de este coeficiente no se puede demostrar que un modelo resulta mejor que el otro, se considerará como el más adecuado, el más sencillo.

La variable objetivo corresponde a: **PRESENTA_DEMORA (dicotómica)**

Las variables independientes serán:

- **Liquidación:** 1 para Si, 0 para No y No definido
- **Obligaciones Ambientales:** 1 para Si, 0 para No.
- **Es post conflicto:** 1 para Si, 0 para No

Estado Contrato: • Borrado • Cerrado • En aprobación • En ejecución • Modificado • Terminado • Activo • Prorrogado • Cedido • Enviado a proveedor.

Indicadores de transparencia:

- Visibilidad.
- Divulgación de la información pública.
- Condiciones institucionales de la información pública.
- dummy información pública.
- dummy visibilidad.
- ITB.

Entrenamiento del modelo

Debido a que la muestra esta desbalanceada en las categorías de la variable respuesta (90% para el NO, 10% para el SI) y debido a las implicaciones que tiene esto en el entrenamiento del modelo (modelo aprende más de la categorías más grande) se decide realizar un muestreo aleatorio simple sobre la categoría más grande, de tal manera que quedaran 15.515 contratos en cada una de a las categorías, posteriormente se selecciona muestras de entrenamiento mediante muestreos estratificado MAS (Muestreo aleatorio simple) del 90% para en entrenamiento y del 10% para la validación.

PRESENTA_MORA	CONTRATOS	EQUILIBRIO DE LAS CATEGORIAS	MUESTRA DE ENTRENAMIENTO (90%)	MUESTRA DE VALIDACIÓN (10%)
NO	147.651	15.515	13.963	1.552
SI	15.515	15.515	13.963	1.552
TOTAL	163.166	31.030	27.926	3.104

Tabla 2. Muestras de entrenamiento y validación.

A continuación, se evidencia los parámetros y estadísticas asociados al mismo del modelo de regresión logística, se muestra en la tabla 3 el modelo final luego del proceso de selección de variables con el fin de obtener el modelo más parsimonioso con la información disponible.

Variable	Coeff.	Std. Error	z-score	p-valor	Odds
Estado Contrato=Modificado	- 1,323	0,03	-46,5	-	0,266
Estado Contrato=Borrador	- 0,470	0,08	-5,8	-	0,625
Estado Contrato=En aprobación	- 0,467	0,09	-5,3	-	0,627
Estado Contrato=terminado	- 0,609	0,05	-12,4	-	0,544
Estado Contrato=cedido	- 0,537	0,10	-5,3	-	0,585
Es Post Conflicto	0,320	0,19	16,7	0,095	1,377
Institucionalidad	0,691	0,14	4,9	-	1,995
Control y Sanción	- 0,910	0,13	-6,9	-	0,402
Divulgación de información pública	- 0,634	0,11	-5,6	-	0,530
ITB_DUMMY	- 0,521	0,05	-10,2	-	0,594
VIS_DUMMY	- 0,425	0,05	-8,9	-	0,654

Tabla 3. Estadísticas modelo de regresión logística (Modelo parsimonioso)

Los coeficientes los interpretaremos de la siguiente manera:

El modelo de regresión logística con una sola variable explicativa este dado por:

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}}$$

De donde se obtiene la siguiente expresión equivalente:

$$\ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x.$$

Donde

$$\pi(x_i) = P(Y_i = 1|x_i).$$

Ahora si asumimos dos contratos i, j distintos que asumen valores de x con diferencia de una unidad ($x^* - x = 1$) tenemos que:

$$\frac{P(y_j = 1|x)}{P(y_j = 0|x)} = e^{\beta_1} \frac{P(y_i = 1|x^*)}{P(y_i = 0|x^*)}$$

Es decir e^{β_1} me indica que tantas veces cambia la probabilidad de que $y=1$ con respecto a la probabilidad que $y=0$ cuando x aumenta una unidad.

Por lo tanto, podemos concluir que:

- Si son positivos indica asociación positiva con la variable respuesta así por ejemplo Es Post Conflicto tiene un coeficiente positivo, lo que indica que si el contrato es post conflicto (toma el valor de 1) aumenta en 1,3 veces la probabilidad de que el contrato caiga en mora con respecto a un contrato que no es post conflicto.
- Ahora si son negativos indica asociación negativa con la variable respuesta, es decir por ejemplo si la divulgación de información pública aumenta en una unidad la probabilidad de caer en demora con respecto a no caer disminuye 0,53.
- Si aumenta la capacidad para generar acciones de control y sanción la probabilidad de caer en demora con respecto a no caer disminuye 0,4.
- En general si el ITB es mayor a 62 la probabilidad de caer en demora con respecto a no caer disminuye 0,6 veces.
- Si la visibilidad es superior a 70 puntos la probabilidad de caer en demora con respecto a no caer disminuye en 0,65 veces.
- Existen algunos coeficientes que contradicen la intuición y es necesario profundizar en estos resultados. Estos están resaltados con rojo. En este caso la institucionalidad presenta una relación no intuitiva, hay que evaluar la posibilidad de que por posibles burocracia dentro de las entidades impida tener un control adecuado de los contratos.

5. Evaluación

6.1. Resultados y análisis

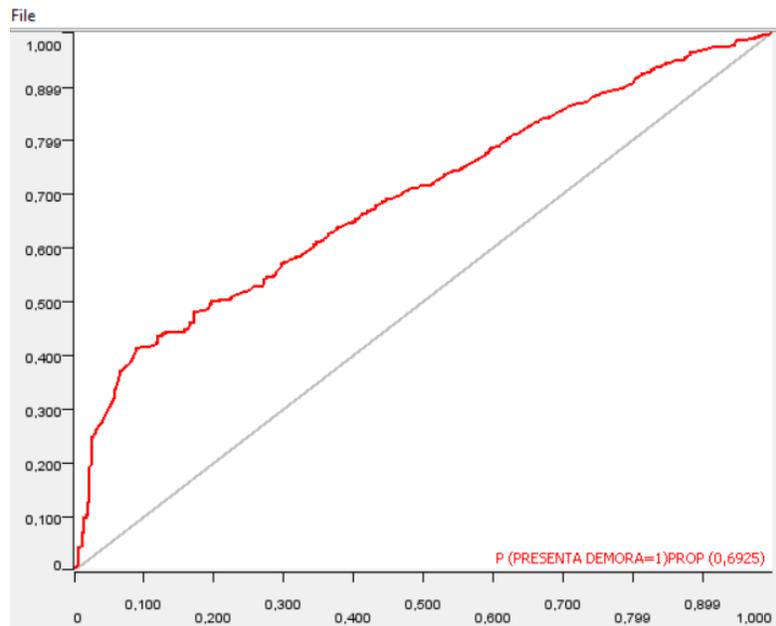


Ilustración 10. Curva COR muestra de entrenamiento

Como se puede evidenciar en el gráfico que corresponde a la curva COR cuya área bajo la curva es de 0,7 aproximadamente, este resultado considera una discriminación aceptable y por encima de la realizada por un clasificador ingenuo.

Ahora se explora la tabla de confusión para la muestra de validación, los resultados muestran un accuracy del 63% y una capacidad de predecir correctamente contratos en mora del 63,2%.

DEMORA/ESTIMACIÓN	1	0	TOTAL	Sensibilidad
1	951	600	1.551	61%
0	553	999	1.552	64%
TOTAL	1.504	1.599	3.103	Accuracy
Especificidad	63,2%	62,48%		63%

Tabla 4. Matrix de confusión muestra de comprobación

Se realizo también una validación cruzada con 100 muestras aleatorias sobre la categoría 1 y se calculo de nuevo el accuracy lo cual arrojó la siguiente distribución:

	Min	P10	P25	P50	P75	P90	MAX
ACURRACY	57,0%	60,0%	64,0%	65,0%	72,0%	75,0%	82,0%

Se puede evidenciar buenos resultados en la validación cruzada, el 80% de las simulaciones obtuvieron un accuracy del 60-75%.

6. Conclusiones

- La necesidad de mejorar la calidad de la información de SECOP I y SECOP II es necesaria, tanto para temas de transparencia como para poder brindar insumos de calidad para la aplicación de Analítica Avanzada que permita darle valor y conocimiento a la sociedad sobre la contratación pública.
- Existe evidencia estadística que los indicadores de transparencia tienen una relación a la hora de explicar la caída de un contrato en demoras, esto es de gran importancia debido a que parte de la variación de las demoras en los contratos puede ser explicada por temas de transparencia en las entidades, no obstante, no son las únicas que logran explicar este fenómeno.
- Bajo los resultados de ajuste se puede evidenciar que efectivamente los factores evaluados especialmente el indicador de transparencia (ITB) son factores que llegan a explicar las demoras en la ejecución de la contratación pública, pero es necesario explorar diferente información exógena no observable en este trabajo para poder tener mejores métricas de ajuste y un modelo más explicativo.

Referencias

Ministerio de tecnologías de la información artículo: El futuro digital es de toda la nueva política TIC (2018). Disponible en: <https://www.mintic.gov.co/portal/inicio/Sala-de-Prensa/Noticias/79186:El-futuro-digital-es-de-todos-la-nueva-politica-TIC>

Alianza CAOBA y ADACOP. Disponible en: <http://alianzacaoba.co/que-es-caoba/>, <http://alianzacaoba.co/adacop/>

OKFN, 2016

Indicador de transparencia del gobierno Nacional. Disponible en: <https://indicedetransparencia.org.co/QueEs>

Datos Abiertos 2018. Disponibles en: <https://www.datos.gov.co/Gastos-Gubernamentales/SECOP-II-Contratos/gnxj-bape>

Documento CONPES: Política Nacional de explotación de datos (BIG DATA). Disponible en: <https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3920.pdf>

ANEXO 2

CARTA DE AUTORIZACIÓN DE LOS AUTORES (Licencia de uso)

Bogotá, D.C., _____ 04 de agosto de 2020 _____

Señores
Biblioteca Alfonso Borrero Cabal S.J.
Pontificia Universidad Javeriana
Ciudad

Los suscritos:

_____ **Oswaldo Martinez Marin** _____, con C.C. No 1014179705
_____, con C.C. No _____
_____, con C.C. No _____

En mi (nuestra) calidad de autor (es) exclusivo (s) de la obra titulada:

MODELO ANALITICO PARA ADACOP

_____ (por favor señale con una “x” las opciones que apliquen)
Tesis doctoral Trabajo de grado Premio o distinción: Si No
cual: _____

presentado y aprobado en el año 2019, por medio del presente escrito autorizo (autorizamos) a la Pontificia Universidad Javeriana para que, en desarrollo de la presente licencia de uso parcial, pueda ejercer sobre mi (nuestra) obra las atribuciones que se indican a continuación, teniendo en cuenta que en cualquier caso, la finalidad perseguida será facilitar, difundir y promover el aprendizaje, la enseñanza y la investigación.

En consecuencia, las atribuciones de usos temporales y parciales que por virtud de la presente licencia se autorizan a la Pontificia Universidad Javeriana, a los usuarios de la Biblioteca Alfonso Borrero Cabal S.J., así como a los usuarios de las redes, bases de datos y demás sitios web con los que la Universidad tenga perfeccionado un convenio, son:

AUTORIZO (AUTORIZAMOS)	SI	NO
1. La conservación de los ejemplares necesarios en la sala de tesis y trabajos de grado de la Biblioteca.	X	
2. La consulta física (sólo en las instalaciones de la Biblioteca)	X	
3. La consulta electrónica - on line (a través del catálogo Biblos y el Repositorio Institucional)	X	
4. La reproducción por cualquier formato conocido o por conocer	X	
5. La comunicación pública por cualquier procedimiento o medio físico o electrónico, así como su puesta a disposición en Internet	X	
6. La inclusión en bases de datos y en sitios web sean éstos onerosos o gratuitos, existiendo con ellos previo convenio perfeccionado con la Pontificia Universidad Javeriana para efectos de satisfacer los fines previstos. En este evento, tales sitios y sus usuarios tendrán las mismas facultades que las aquí concedidas con las mismas limitaciones y condiciones	X	

De acuerdo con la naturaleza del uso concedido, la presente licencia parcial se otorga a título gratuito por el máximo tiempo legal colombiano, con el propósito de que en dicho lapso mi (nuestra) obra sea explotada en las condiciones aquí estipuladas y para los fines indicados, respetando siempre la titularidad de los derechos patrimoniales y morales correspondientes, de

acuerdo con los usos honrados, de manera proporcional y justificada a la finalidad perseguida, sin ánimo de lucro ni de comercialización.

De manera complementaria, garantizo (garantizamos) en mi (nuestra) calidad de estudiante (s) y por ende autor (es) exclusivo (s), que la Tesis o Trabajo de Grado en cuestión, es producto de mi (nuestra) plena autoría, de mi (nuestro) esfuerzo personal intelectual, como consecuencia de mi (nuestra) creación original particular y, por tanto, soy (somos) el (los) único (s) titular (es) de la misma. Además, aseguro (aseguramos) que no contiene citas, ni transcripciones de otras obras protegidas, por fuera de los límites autorizados por la ley, según los usos honrados, y en proporción a los fines previstos; ni tampoco contempla declaraciones difamatorias contra terceros; respetando el derecho a la imagen, intimidad, buen nombre y demás derechos constitucionales. Adicionalmente, manifiesto (manifestamos) que no se incluyeron expresiones contrarias al orden público ni a las buenas costumbres. En consecuencia, la responsabilidad directa en la elaboración, presentación, investigación y, en general, contenidos de la Tesis o Trabajo de Grado es de mi (nuestro) competencia exclusiva, eximiendo de toda responsabilidad a la Pontificia Universidad Javeriana por tales aspectos.

Sin perjuicio de los usos y atribuciones otorgadas en virtud de este documento, continuare (continuaremos) conservando los correspondientes derechos patrimoniales sin modificación o restricción alguna, puesto que de acuerdo con la legislación colombiana aplicable, el presente es un acuerdo jurídico que en ningún caso conlleva la enajenación de los derechos patrimoniales derivados del régimen del Derecho de Autor.

De conformidad con lo establecido en el artículo 30 de la Ley 23 de 1982 y el artículo 11 de la Decisión Andina 351 de 1993, “*Los derechos morales sobre el trabajo son propiedad de los autores*”, los cuales son irrenunciables, imprescriptibles, inembargables e inalienables. En consecuencia, la Pontificia Universidad Javeriana está en la obligación de RESPETARLOS Y HACERLOS RESPETAR, para lo cual tomará las medidas correspondientes para garantizar su observancia.

NOTA: Información Confidencial:

Esta Tesis o Trabajo de Grado contiene información privilegiada, estratégica, secreta, confidencial y demás similar, o hace parte de una investigación que se adelanta y cuyos resultados finales no se han publicado. Si No

En caso afirmativo expresamente indicaré (indicaremos), en carta adjunta, tal situación con el fin de que se mantenga la restricción de acceso.

NOMBRE COMPLETO	No. del documento de identidad	FIRMA
Oswaldo Martinez Marin	1014179705	<i>Oswaldo Martinez Marin</i>

FACULTAD: Ingeniería

PROGRAMA ACADÉMICO: Maestría en Analítica para la inteligencia de negocios

ANEXO 3
BIBLIOTECA ALFONSO BORRERO CABAL, S.J.
DESCRIPCIÓN DE LA TESIS O DEL TRABAJO DE GRADO
FORMULARIO

TÍTULO COMPLETO DE LA TESIS DOCTORAL O TRABAJO DE GRADO						
MODELO ANALITICO PARA ADACOP						
SUBTÍTULO, SI LO TIENE						
AUTOR O AUTORES						
Apellidos Completos			Nombres Completos			
Martinez Marin			Oswaldo			
DIRECTOR (ES) TESIS O DEL TRABAJO DE GRADO						
Apellidos Completos			Nombres Completos			
Pulido Moreno			Luis Manuel			
Sierra Múnera			Alejandro			
FACULTAD						
ingeniería						
PROGRAMA ACADÉMICO						
Tipo de programa (seleccione con "x")						
Pregrado	Especialización	Maestría	Doctorado			
		x				
Nombre del programa académico						
Maestría en Analítica para la inteligencia de Negocios						
Nombres y apellidos del director del programa académico						
Jorge Alvarado						
TRABAJO PARA OPTAR AL TÍTULO DE:						
Magister Analítica para la inteligencia de Negocios						
PREMIO O DISTINCIÓN (En caso de ser LAUREADAS o tener una mención especial):						
CIUDAD		AÑO DE PRESENTACIÓN DE LA TESIS O DEL TRABAJO DE GRADO			NÚMERO DE PÁGINAS	
Bogotá		2019			31	
TIPO DE ILUSTRACIONES (seleccione con "x")						
Dibujos	Pinturas	Tablas, gráficos y diagramas	Planos	Mapas	Fotografías	Partituras
		x				
SOFTWARE REQUERIDO O ESPECIALIZADO PARA LA LECTURA DEL DOCUMENTO						
<p>Nota: En caso de que el software (programa especializado requerido) no se encuentre licenciado por la Universidad a través de la Biblioteca (previa consulta al estudiante), el texto de la Tesis o Trabajo de Grado quedará solamente en formato PDF.</p>						

MATERIAL ACOMPAÑANTE					
TIPO	DURACIÓN (minutos)	CANTIDAD	FORMATO		
			CD	DVD	Otro ¿Cuál?
Vídeo					
Audio					
Multimedia					
Producción electrónica					
Otro Cuál?					
DESCRIPTORES O PALABRAS CLAVE EN ESPAÑOL E INGLÉS					
Son los términos que definen los temas que identifican el contenido. <i>(En caso de duda para designar estos descriptores, se recomienda consultar con la Sección de Desarrollo de Colecciones de la Biblioteca Alfonso Borrero Cabal S.J en el correo biblioteca@javeriana.edu.co, donde se les orientará).</i>					
ESPAÑOL			INGLÉS		
CAOBA			CAOBA		
Analítica			analytics		
SECOP II			SECOP II		
Contratación pública			public procurement		
RESUMEN DEL CONTENIDO EN ESPAÑOL E INGLÉS (Máximo 250 palabras - 1530 caracteres)					
<p>Uno de los objetivos estratégicos de CAOBA es impulsar la gestión del conocimiento en analítica. Parte fundamental de la gestión del conocimiento es su efectiva trasferencia, esta se ve fortalecido en el trabajo conjunto entre CAOBA y el Observatorio fiscal de la Pontificia Universidad Javeriana; identificando que herramientas analíticas (modelos) pueden apoyar la transparencia en la contratación pública del estado colombiano aportando a todos los actores involucrados (ciudadanos, entidades públicas, entro otros) un mayor entendimiento del proceso de contratación estatal, robusteciendo la actual plataforma ADACOP. Este trabajo plantea la manera de poder identificar características de las entidades del estado que me permitan establecer un nivel de riesgo de caer en demoras en la contratación o sobrecostos, esto se logra integrando la información publica del estado en temas de transparencia de las entidades y la información disponible de la contratación pública y las metodologías estadísticas y de machine learning.</p>					

One of CAOBA's strategic objectives is to promote analytical knowledge management. A fundamental part of knowledge management is its effective transfer, this is strengthened in the joint work between CAOBA and the Fiscal Observatory of the Pontificia Universidad Javeriana; identifying which analytical tools (models) can support transparency in public procurement in the Colombian state, providing all the actors involved (citizens, public entities, among others) with a greater understanding of the state procurement process, strengthening the current ADACOP platform. This work proposes the way to identify characteristics of state entities that allow me to establish a level of risk of falling into delays in contracting or cost overruns, this is achieved by integrating the state's public information on issues of transparency of entities and the Information available on public procurement and statistical and machine learning methodologies.