



Mejorar la contratación pública: Herramientas informáticas innovadoras basadas en datos e inteligencia artificial

Manuel J. García Rodríguez

Instituto Andaluz de Administración Pública | Premios Blas Infante XIX Edición

**MEJORAR LA CONTRATACIÓN PÚBLICA: HERRAMIENTAS
INFORMÁTICAS INNOVADORAS BASADAS EN DATOS
E INTELIGENCIA ARTIFICIAL**

Trabajo galardonado con el Premio Blas Infante de Estudio e Investigación en el ámbito de la Comunidad Autónoma, en su XIX edición en la modalidad B. “Innovación en la Administración Pública”, incluyendo estudios de ámbito internacional, que reflexionen sobre nuevas maneras de abordar los distintos ámbitos de la acción pública o supongan actuaciones innovadoras en sus respectivas áreas, relativas a su organización, funciones, modernización de la gestión del servicio público, innovación en servicios y procesos y aplicación de nuevas tecnologías.

**MEJORAR LA CONTRATACIÓN PÚBLICA: HERRAMIENTAS
INFORMÁTICAS INNOVADORAS BASADAS EN DATOS
E INTELIGENCIA ARTIFICIAL**

Manuel J. García Rodríguez

Instituto Andaluz de Administración Pública
Sevilla 2023

Mejorar la contratación pública: Herramientas informáticas innovadoras basadas en datos e inteligencia artificial / Manuel José García Rodríguez

Sevilla: Instituto Andaluz de Administración Pública, 2023
107 p. ; 24 cm. - (Estudios).

Trabajo galardonado con el Premio Blas Infante de Estudio e Investigación en el ámbito de la Comunidad Autónoma, en su XIX edición en la modalidad B. "Innovación en la Administración Pública", incluyendo estudios de ámbito internacional, que reflexionen sobre nuevas maneras de abordar los distintos ámbitos de la acción pública o supongan actuaciones innovadoras en sus respectivas áreas, relativas a su organización, funciones, modernización de la gestión del servicio público, innovación en servicios y procesos y aplicación de nuevas tecnologías.

D.L.: SE 702-2024

ISBN (edición impresa) 978-84-8333-724-0

1. Contratos administrativos 2. Innovaciones tecnológicas. 3. Inteligencia Artificial

RESERVADOS TODOS LOS DERECHOS. NO ESTÁ PERMITIDA LA REPRODUCCIÓN TOTAL O PARCIAL EN NINGÚN TIPO DE SOPORTE SIN PERMISO PREVIO Y POR ESCRITO DEL TITULAR DEL COPYRIGHT

MEJORAR LA CONTRATACIÓN PÚBLICA: HERRAMIENTAS INFORMÁTICAS INNOVADORAS BASADAS EN DATOS E INTELIGENCIA ARTIFICIAL

AUTOR: Manuel José García Rodríguez

© INSTITUTO ANDALUZ DE ADMINISTRACIÓN PÚBLICA

Edita: Instituto Andaluz de Administración Pública

Coordinación y corrección: Francisco Javier Domínguez Murillo

Diseño, maquetación e impresión: Imprenta Flores

ISBN: 978-84-8333-724-0

Depósito Legal: SE 702-2024

ÍNDICE

ÍNDICE DE FIGURAS	11
ÍNDICE DE TABLAS	13
GLOSARIO	15
RESUMEN	17
CAPÍTULO PRIMERO. INTRODUCCIÓN	19
1.1. MOTIVACIÓN	19
1.2. ANTECEDENTES	20
1.3. ESTRUCTURA DEL DOCUMENTO	22
1.4. PUBLICACIONES REALIZADAS DURANTE LA TESIS	24
CAPÍTULO SEGUNDO. RETOS Y OBJETIVOS	27
2.1. RETOS ACTUALES EN LA CONTRATACIÓN PÚBLICA	27
2.2. ORGANISMOS PÚBLICOS ESPAÑOLES RELACIONADOS CON CONTRATACIÓN	30
2.3. TRANSPARENCIA Y DATOS EN ABIERTO DE CONTRATACIÓN EN ESPAÑA	32

2.4. INICIATIVAS DE LA COMISIÓN EUROPEA ASOCIADAS A DATOS DE CONTRATACIÓN	33
2.5. OBJETIVOS Y METODOLOGÍA DE LA INVESTIGACIÓN	36
CAPÍTULO TERCERO. FUNDAMENTOS DE LA INVESTIGACIÓN.....	39
3.1. LA CONTRATACIÓN PÚBLICA	39
3.2. LEGISLACIÓN EN LA CONTRATACIÓN PÚBLICA.....	40
3.3. TECNOLOGÍAS EN LA CONTRATACIÓN PÚBLICA.....	43
3.4. APLICACIONES EN LA CONTRATACIÓN PÚBLICA	45
3.5. LA PLATAFORMA DE CONTRATACIÓN DEL SECTOR PÚBLICO (PLACSP)	49
3.5.1. Órganos de contratación que publican en PLACSP.....	49
3.5.2. Origen de los datos de PLACSP.....	50
3.5.3. Formato y calidad de los datos en abierto de PLACSP	51
3.5.4. Programa OpenPLACSP para analizar los datos en abierto.....	53
3.6. LITERATURA ACADÉMICA SOBRE LA CONTRATACIÓN PÚBLICA	53
3.7. MÉTRICAS DE EVALUACIÓN PARA LOS ALGORITMOS DE ML.....	57
3.7.1. Métricas de evaluación para problemas de regresión	58
3.7.2. Métricas de evaluación para problemas de clasificación.....	59
CAPÍTULO CUARTO. ARTÍCULOS DE LA INVESTIGACIÓN	63
4.1. INTRODUCCIÓN.....	63
4.2. LICITACIONES EN ESPAÑA: REGULACIÓN, ANÁLISIS DE DATOS Y ESTIMADOR DEL IMPORTE DE ADJUDICACIÓN USANDO ML	65
4.3. ESTIMADOR DEL IMPORTE DE ADJUDICACIÓN DE LICITACIONES USANDO ML: CASO DE ESTUDIO CON LICITACIONES DE ESPAÑA.....	69

4.4. RECOMENDADOR DE LICITADORES USANDO ML: ANÁLISIS DE DATOS, ALGORITMO Y CASO DE ESTUDIO CON LICITACIONES DE ESPAÑA	71
4.5. DETECCIÓN DE COLUSIÓN EN LICITACIONES APLICANDO ALGORITMOS DE ML	75
4.6. APLICACIÓN INFORMÁTICA PARA DETECTAR LICITACIONES IRREGULARES	80
CAPÍTULO QUINTO. DISCUSIÓN DE LOS RESULTADOS	83
CAPÍTULO SEXTO. CONCLUSIONES	91
6.1. CONCLUSIONES	91
6.2. LÍNEAS DE INVESTIGACIÓN FUTURAS	93
BIBLIOGRAFÍA	95

ÍNDICE DE FIGURAS

2.1. Ejemplo de supervisión de la contratación de España. Proyecto Open Tender	29
2.2. Mapa de los organismos públicos españoles ligados a la contratación	31
2.3. Elementos para tener un espacio de datos de contratación pública europeo unificado.....	34
3.1. Funcionamiento de un algoritmo de ML: entrenamiento (a) y uso (b). Fuente [46]	44
3.2. Diagrama de posibles aplicaciones para la contratación pública.....	46
3.3. Web de la Plataforma de Contratación del Sector Público (PLACSP)	49
3.4. Programa OpenPLACSP para extraer licitaciones de Plataforma de Contratación del Sector Público (PLACSP)	54
3.5. Elementos que componen un diagrama de cajas (boxplot)	57
3.6. Matriz de confusión para la clasificación binaria.....	61
4.1. Flujograma del análisis de datos de contratación y el estimador del precio de adjudicación. A la derecha, las librerías software utilizadas	66

4.2. En la gráfica superior se compara el importe de licitación (eje x) frente al importe de adjudicación (eje y). En la inferior, la predicción del importe de adjudicación (eje x) frente al importe de adjudicación (eje y). Datos de España	68
4.3. Arquitectura de las capas y nodos de las redes neuronales artificiales (ANN) utilizadas	70
4.4. Funcionamiento del recomendador de licitadores (buscador de empresas) mediante dos capturas de pantalla de la aplicación web	72
4.5. Flujograma del funcionamiento del recomendador de licitadores (buscador de empresas) para una nueva licitación de entrada	73
4.6. Flujograma para crear el algoritmo de ML recomendador de empresas.....	74
4.7. Flujograma para la detección de colusión mediante ML.....	77
4.8. Gráficas del error (precision, recall y F1 score) para el dataset conjunto. A la izquierda para el setting 3 (campos comunes) y a la derecha para el setting 4 (campos comunes + screens)	80
4.9. Aplicación informática para detectar licitaciones irregulares recogida en la noticia de El País publicada el 10/12/2021.....	82
5.1. Diagrama de cajas (boxplot) del error porcentual absoluto (APE) entre el importe de adjudicación y la predicción (gris) y el APE entre el importe de adjudicación y licitación (azul). Agrupados por CPV para los datos de España. Fuente: elaboración propia en [10]	85
5.2. Histograma que representa el número de empresas (eje y) que han ganado el mismo número de licitaciones (eje x). Se ha dividido el gráfico en dos para una mejor visualización. Fuente: elaboración propia en [12]	86
5.3. Diferencia porcentual entre el importe de licitación y adjudicación (baja económica) (eje y) según el número de ofertas recibidas en la licitación (eje x). La gráfica superior es la baja mediana (elaboración propia en [12]) y la gráfica inferior es la baja media (elaborada por la Oficina Independiente de Regulación y Supervisión de la Contratación (OIReScon) en [21]).....	88
5.4. Porcentaje de licitaciones que han tenido una única oferta, dividido por países y años. Fuente: Comisión Europea (CE).....	90

ÍNDICE DE TABLAS

1.1. Revistas de los artículos publicados para la Tesis	25
3.1. Legislación sobre contratación pública y el uso de datos.....	41
3.2. Recopilación de la literatura relacionada con los datos en abierto y la calidad del dato, la innovación y gestión en la contratación.....	55
3.3. Recopilación de la literatura relacionada con el forecasting en la contratación y la colusión y corrupción.....	55-56
4.1. Datasets utilizados y métricas de error del estimador del precio de adjudicación.....	67
4.2. Métricas de error del estimador del precio de adjudicación para los algoritmos de ML.....	71
4.3. Métricas de error del recomendador de licitadores para distintas configuraciones en dos escenarios diferentes	75
4.4. Descripción de los datasets (licitaciones competitivas y colusivas) de Brasil, Italia, Japón, Suiza-Ticino, Suiza-SG&GR y USA	76
4.5. Resumen de los resultados de detección de colusión para los diferentes datasets (Brasil, Italia, Japón, Suiza-Ticino, Suiza-SG&GR, EE.UU. y en conjunto).....	78
5.1. Diferencia entre el importe de licitación y adjudicación (baja económica) para diferentes grupos de ofertas recibidas. Fuente: elaboración propia en [10]	87

GLOSARIO

AA.PP.	Administraciones Públicas
AGE	Administración General del Estado
AIReF	Autoridad Independiente de Responsabilidad Fiscal
CC.AA.	Comunidades Autónomas
CE	Comisión Europea
CNMC	Comisión Nacional de los Mercados y la Competencia
CODICE	Componentes y Documentos Interoperables para la Contratación Electrónica
CPV	Common Procurement Vocabulary
DGRCC	Dirección General de Racionalización y Centralización de la Contratación
DOUE	Diario Oficial de la Unión Europea
IA	Inteligencia Artificial
IGAE	Intervención General de la Administración del Estado

LCSP	Ley 9/2017 de Contratos del Sector Público
MAPE	Mean Absolute Percentage Error
MdAPE	Median Absolute Percentage Error
ML	Machine Learning
OIREscon	Oficina Independiente de Regulación y Supervisión de la Contratación
PLACSP	Plataforma de Contratación del Sector Público
TED	Tenders Electronic Daily
UE	Unión Europea

RESUMEN

La contratación pública es el gasto público para proveer de bienes, servicios o trabajos a una entidad pública. Es un campo de gran importancia por representar un porcentaje significativo del gasto sobre el PIB de los Estados, un 16% según algunas estimaciones oficiales. Sin embargo, es un campo muy poco estudiado por los investigadores del ámbito científico-tecnológico porque hasta hace pocos años no se disponían de datos de contratos públicos (licitaciones) con la información estructurada y accesible para su descarga masiva por cualquier ciudadano.

Este trabajo tiene como objetivo explicar la Tesis Doctoral realizada por el autor (un ingeniero) a un lector no técnico para difundir la investigación realizada: aplicar la ciencia de datos a la contratación pública. En particular, se analizaron los datos de las licitaciones, tanto de España como de la Unión Europea, y se abordaron algunos de los problemas existentes en este campo aplicando los algoritmos de aprendizaje automático, Machine Learning (ML), que se encuadran dentro de la Inteligencia Artificial (IA). Se explicarán las herramientas informáticas desarrolladas: un estimador del importe de adjudicación en licitaciones públicas, un buscador/recomendador de empresas para una licitación dada, una herramienta que detecta licitaciones irregulares gracias a los metadatos de los pliegos o un detector de colusión (prácticas anticompetitivas que hacen varias empresas formando un cártel). Por tanto, es una investigación innovadora en el campo académico, en la Administración y en el sector privado.

Con este trabajo se pretende tender un puente de unión entre dos ámbitos bastante distantes: la contratación pública y la ciencia de datos. El primero lo lideran expertos de las ciencias sociales y el segundo son expertos del área científico-tecnológica. Ambos

colectivos se verán beneficiados si trabajan de manera conjunta, colaborando para conseguir una contratación más eficiente, segura y digitalizada.

En conclusión, este trabajo recoge herramientas innovadoras, prácticas y reales, que aportan valor a las AA.PP. que gestionan, controlan y auditan la contratación.

Además, los operadores económicos que licitan también obtendrían ventajas y beneficios al usarlas. La originalidad, innovación y utilidad para resolver problemas reales utilizando las herramientas explicadas en este trabajo es notoria. Por ejemplo, la CNMC (Comisión Nacional de los Mercados de la Competencia) utiliza la herramienta de colusión para luchar contra los cárteles. Otro ejemplo, se han detectado una veintena de licitaciones irregulares en España (incluyendo varias en Andalucía) dando lugar a noticias publicadas en periódicos nacionales y que en la actualidad los organismos públicos competentes han admitido a trámite para investigarlas y enjuiciarlas.

CAPÍTULO PRIMERO. INTRODUCCIÓN

1.1. MOTIVACIÓN

La contratación pública es la denominación común del gasto público para proveer de bienes, servicios o trabajos a una entidad pública [1]. La contratación pública es un intensivo y complejo proceso que consume cuantiosos recursos. Por ejemplo, la Unión Europea (UE) gasta alrededor del 16% de su PIB en contratos públicos [2]. Por tanto, los mayores adjudicadores de contratos de un país, tanto por número como por importe, son los organismos públicos y cualquier pequeña eficiencia que se logre en sus procesos supondrá un ahorro significativo.

Esta Tesis aplica la ciencia de datos (data science) a la contratación pública, combinación innovadora tanto en el campo académico como en el sector privado o en las Administraciones Públicas (AA.PP.) ¿Qué es la ciencia de datos? Porque toda ciencia involucra datos. No es concepto fácil de definir y más si cabe en los últimos años donde se ha vuelto omnipresente. En la actualidad, es un campo multidisciplinar que utiliza métodos científicos (procesos, algoritmos, etc.) para extraer conocimientos de los datos mediante la informática. La ciencia de datos se puede definir como la unión de 3 campos: conocimiento matemático y estadístico, conocimientos y habilidades de programación y el conocimiento experto del objeto de estudio (en el caso que nos ocupa, la contratación) [3]. En particular, en esta Tesis se estudiarán los datos de los contratos públicos (licitaciones) y se abordarán algunos de los problemas que hay en este campo aplicando la ciencia de datos y, en particular, los algoritmos de aprendizaje automático, Machine Learning (ML) (se explicará en el capítulo 3), que se encuadran dentro de la IA.

Siempre que nos refiramos a la contratación, se sobrentenderá que es contratación pública. Esta Tesis no tiene por objeto estudiar las particularidades de la contratación

privada, esto es, la realizada por empresas dentro de sus actividades económicas. Sin embargo, ambos tipos de contratación comparten muchas características [4] porque solamente se diferencian en la naturaleza, pública o privada, de la entidad contratante. La gran mayoría de problemas enunciados, metodologías propuestas, aplicaciones desarrolladas y conclusiones plasmadas en esta Tesis se pueden adaptar y aplicar a la contratación privada.

La contratación pública no ha sido ajena a la ciencia de datos o a la IA. Los estudios y opiniones sobre su aplicabilidad a la contratación son cada vez más comunes. Sin embargo, las aplicaciones de este tipo de tecnologías en la práctica administrativa de la compra pública son, por el momento, escasas, dispersas y veladas [5]. Es decir, hay una gran carencia de herramientas innovadoras que utilicen la IA y esta Tesis trata de paliarlo usando uno de sus subcampos más relevantes: el ML. En otros ámbitos de las AA.PP., no limitándose a la contratación, cada vez vanteniendo más ejemplos exitosos que explotan los datos que manejan gracias a la IA [6].

Esta Tesis trata de ser un puente entre dos ámbitos bastante distantes entre sí hasta la fecha: la contratación pública y la ciencia de datos. El primero está liderado por expertos del ámbito de las ciencias sociales (juristas, economistas, politólogos, etc.) y el segundo por expertos del ámbito científico-tecnológico (ingenieros, matemáticos, informáticos, etc.). Ambos colectivos se verán beneficiados y recompensados si trabajan de manera conjunta, colaborando para conseguir una contratación más eficiente, segura y digitalizada. Por tanto, se puede denominar a esta Tesis de frontera, donde el autor hará especial esfuerzo en que ambos grupos de personas entiendan el campo de conocimiento del otro. Es decir, que la Tesis sea comprensible para todos y sirva de ejemplo para futuros proyectos transversales y multidisciplinares.

En los siguientes apartados se resumirá los antecedentes de la ciencia de datos y la contratación, la estructura del documento (6 capítulos) y las publicaciones académicas técnicas que se explicarán en este trabajo.

1.2. ANTECEDENTES

En este apartado se va a hacer un breve resumen de la **evolución histórica de la ciencia de datos y la contratación pública** para tener una visión amplia de ambos campos. Hay dos diferencias sustanciales entre ambos, una de carácter geográfico y otra temporal. La ciencia de datos se ha desarrollado a nivel mundial, cooperando personas de distintos países, a partir del último tercio del siglo XX y de manera muy intensa en los últimos años hasta la actualidad. Sin embargo, la contratación pública tiene una evolución

propia y diferencia para cada país, por formar parte de su legislación (derecho administrativo), que en el caso español nace ya en el siglo XIX.

El nacimiento de cualquier campo del saber es muy difícil de ubicar y fechar y el caso de la ciencia de datos no es diferente. Un hecho primigenio y relevante tuvo lugar en 1962 cuando el profesor John W. Tukey¹ (Univ. de Princeton) escribió “*The Future of Data Analysis*” [7] donde esboza una nueva ciencia sobre el aprendizaje de los datos, instando a los académicos a reducir su enfoque en la teoría estadística y a participar en todo el proceso de análisis de datos:

“Durante mucho tiempo pensé que era un estadístico interesado en inferencias de lo particular a lo general. Pero a medida que observé la evolución de las estadísticas matemáticas, tuve motivos para preguntarme y dudar [...] Llegué a sentir que mi interés central está en el análisis de datos. El análisis de datos, y las partes de las estadísticas que se adhieren a él, deben [...] asumir las características de la ciencia en lugar de las matemáticas [...] el análisis de datos es intrínsecamente una ciencia empírica.”

Aquí se menciona la evolución de la estadística matemática como ciencia de datos. Sin embargo, no sería hasta 1974 cuando Peter Naur [8], científico danés pionero en la informática, acuñara el término que actualmente conocemos. El desarrollo de las ciencias de la computación (software), la microelectrónica (hardware) y las redes de telecomunicaciones a finales del siglo XX y comienzos del XXI, gracias a la inversión económica realizada por el sector privado (grandes multinacionales) y público (Universidades y centros tecnológicos de investigación), han conseguido un altísimo nivel de informatización, de digitalización de la sociedad, que ha supuesto una revolución social y económica. Esto ha sentado las bases para que en los últimos 10 años haya habido una explosión de proyectos basados en ciencia de datos en muchos sectores de la economía (industria, banca, telecomunicaciones, energía, transporte, sanidad, etc.) y, algo más tarde, también en las AA.PP.

La evolución histórica de la contratación pública en España se remonta al Real Decreto de 27 febrero de 1852, publicado por Bravo Murillo y que decía “*proyecto de ley de contratos sobre servicios públicos, con el fin de establecer ciertas trabas saludables, evitando los abusos fáciles de cometer en una materia de peligrosos estímulos, y de garantizar la Administración contra los tiros de la maledicencia...*” [9]. Es el nacimiento de la contratación administrativa en España y ya menciona 3 elementos cruciales que aún hoy siguen estando presentes: limitar los abusos fáciles (de autoridades y funcionarios) contra peligrosos estímulos (cohechos, prebendas, sobornos) y evitar los tiros de la maledicencia (desconfianza en los servidores públicos).

¹ Estadístico experto en topología y conocido, entre otras aportaciones, por el cálculo de la transformada rápida de Fourier (FFT) y el diagrama de cajas (boxplot).

Después, se dictaron otras disposiciones de igual rango, hasta que en 1911 se promulgó la Ley de Contabilidad de la Hacienda Pública, en cuyo capítulo V se trataba la contratación. La única forma de adjudicación de los contratos era mediante la subasta (la proposición más económica, sin importar la calidad, plazos u otras características del suministro). Estas disposiciones estuvieron vigentes hasta la Ley de Bases de Contratación del Estado de 1963 y su texto articulado de 1965, finalizándose la regulación de esta materia con la promulgación del Reglamento de Contratación del Estado en 1967. En 1975 se actualizó mediante el Decreto 3410/1975 y, tras la incorporación de España en la Comunidad Económica Europea en 1986, también hubo una serie de adaptaciones a la legislación comunitaria europea. En 1995 se derogan expresamente las leyes anteriores de contratación y se crea la Ley de contratos de las Administraciones Públicas (Ley 13/1995). En los años posteriores se sucederán varias leyes de contratación que buscan incrementar la concurrencia y aumentar la transparencia y objetividad en los procedimientos de adjudicación, así como simplificar en lo posible los procedimientos de contratación, adaptarlos a la digitalización de las AA.PP. La actual ley en vigor es la Ley 9/2017 de Contratos del Sector Público (LCSP) y abre las puertas a poder utilizar los datos de las licitaciones de manera masiva, aplicando la ciencia de datos y las técnicas de ML.

1.3. ESTRUCTURA DEL DOCUMENTO

Se ha hecho un esfuerzo en tratar el objeto de estudio, la contratación pública, no sólo como un mero conjunto de datos en donde aplicar la ciencia de datos, esto es, la experimentación con algoritmos de ML y técnicas cuantitativas, sino en comprender su naturaleza desde distintas perspectivas (legal, de gestión pública, económica, etc.). Con esta visión más holística se puede investigar con conocimiento de causa, aportando soluciones realistas e innovadoras mediante el ML. El presente trabajo se ha estructurado en 6 capítulos, resumidos a continuación.

Capítulo 1: Introducción. Se describe la motivación para haber realizado la investigación, así como los antecedentes y relevancia que tiene la contratación pública.

Capítulo 2: Retos y objetivos. Se explican los retos actuales de la contratación y los organismos públicos con competencias en esta materia. Además, se resumen los conceptos de la transparencia en la contratación, los datos en abierto (open data) y las iniciativas de la CE ligadas a los datos de contratación a nivel europeo, por ser elementos necesarios para desarrollar esta Tesis. Es decir, la investigación se ha basado fundamentalmente en datos públicos, accesibles a cualquier ciudadano. Finalmente, se enumeran los objetivos y metodología que se han marcado en la Tesis.

Capítulo 3: Fundamentos de la investigación. Se describe la contratación desde el punto de vista cuantitativo, legislativo, de tecnologías software y sus posibles aplicaciones (casos de uso). Por ser la principal fuente de datos empleada, se detalla la Plataforma de Contratación del Sector Público (PLACSP): los órganos que publican y adjudican las licitaciones y el origen de los datos así como su formato y calidad. Se hace una recopilación de la literatura utilizada en la contratación para que se tengan referencias del estado del arte actual. Además, se formulan varias métricas de error típicas para evaluar los algoritmos de ML, en particular para los problemas de regresión y clasificación (los tratados en la Tesis).

Capítulo 4: Investigación realizada. Se resumen, explicándose para un lector no técnico, los 4 artículos que forman parte de la Tesis y son el grueso de la investigación (teórica y práctica) efectuada. Además, se describe un programa informático, desarrollado por el autor, que detecta licitaciones irregulares. En particular, ha detectado una veintena de licitaciones potencialmente fraudulentas en España (y también en Andalucía) que suman más de 10 millones de euros.

- “Public Procurement Announcements in Spain: Regulations, Data Analysis, and Award Price Estimator Using Machine Learning.” En castellano es *“Licitaciones en España: regulación, análisis de datos y estimador del importe de adjudicación usando ML”*.
- “Award Price Estimator for Public Procurement Auctions Using Machine Learning Algorithms: Case Study with Tenders from Spain.” En castellano es *“Estimador del importe de adjudicación de licitaciones usando ML: caso de estudio con licitaciones de España”*.
- “Bidders Recommender for Public Procurement Auctions Using Machine Learning: Data Analysis, Algorithm, and Case Study with Tenders from Spain.” En castellano es *“Recomendador de licitadores usando ML: análisis de datos, algoritmo y caso de estudio con licitaciones de España”*.
- “Collusion detection in public procurement auctions with machine learning algorithms.” En castellano es *“Detección de colusión en licitaciones aplicando algoritmos de ML”*.

Capítulo 5: Discusión de los resultados. Análisis de los resultados más relevantes de los anteriores artículos, con una visión de conjunto y transversal. Se analiza primero desde una perspectiva técnica, algorítmica, para evaluar la aplicación de los algoritmos de ML en el campo de la contratación. Después, se analiza desde una perspectiva cuantitativa, económica, para abrir el campo a futuros estudios económicos. Así, esta Tesis se puede tomar como un

ejemplo pionero en la utilización de la ciencia de datos para hacer estudios económicos en la contratación.

Capítulo 6: Conclusiones. Conclusión de la investigación y futuras líneas de trabajo.

1.4. PUBLICACIONES REALIZADAS DURANTE LA TESIS

A continuación, se citan las publicaciones realizadas dentro del marco de la investigación de la Tesis llevada a cabo durante los últimos 5 años. En la Tabla 1.1 se citan los artículos que forman parte del compendio de publicaciones de la Tesis y en qué revistas se publicaron (todas pertenecen al Science Citation Index Expanded, SCIE).

- [10] “Public Procurement Announcements in Spain: Regulations, Data Analysis, and Award Price Estimator Using Machine Learning.”
- [11] “Award Price Estimator for Public Procurement Auctions Using Machine Learning Algorithms: Case Study with Tenders from Spain.”
- [12] “Bidders Recommender for Public Procurement Auctions Using Machine Learning: Data Analysis, Algorithm, and Case Study with Tenders from Spain.”
- [13] “Collusion detection in public procurement auctions with machine learning algorithms.”
- [15] “Spanish Public Procurement: Legislation, open data source and extracting valuable information of procurement announcements.”
- [15] “La contratación pública en España: fuentes de datos, normativa y aplicaciones tecnológicas.”
- [16] “Tecnologías digitales para el control de la contratación pública.”

Además, hasta la fecha actual se ha participado en los siguientes congresos para compartir y debatir las investigaciones llevadas a cabo:

- Comunicación titulada “*La Contratación Pública en España: fuentes de datos, normativa y aplicaciones tecnológicas*” [15] y conferencia impartida en el *Congreso Internacional online: temas clave de la contratación pública*. Organizado por la Univ. de Vigo el día 5/10/2021.
- Comunicación titulada “*Mejorar la gestión y supervisión de la contratación: los algoritmos de IA y los datos en abierto de PLACSP*” y conferencia impartida en el *XI Congreso Internacional sobre contratación pública*. Organizado por la Univ. de Castilla-La Mancha los días 25-26/1/2022 en Cuenca. La comunicación

ha sido seleccionada para publicarse en el libro *Observatorio de los contratos públicos 2022*, editorial Aranzadi (pendiente de publicación).

- Conferencia online titulada “*La implantación de la IA en la contratación pública: un diálogo entre tecnología y derecho*” impartida en el *II Seminario de expertos sobre derecho administrativo e innovación tecnológica*. Organizado por la Univ. de Castilla-La Mancha el 15/3/2022.

Revista	Editor	Artículo	Factor de impacto	Categorías y ranking por factor de impacto
<i>Automation in construction</i>	Elsevier	[13]	7.700	Civil engineering (2020) = 2/137 (Q1); Construction & Building technology (2020) = 3/67 (Q1)
<i>Complexity</i>	Wiley- Hindawi	[10]	2.462	Mathematics, interdisciplinary applications (2019) = 28/106 (Q2); Multidisciplinary sciences (2019) = 31/71 (Q2)
<i>Complexity</i>	Wiley- Hindawi	[12]	2.833	Mathematics, interdisciplinary applications (2020) = 31/108 (Q2); Multidisciplinary sciences (2020) = 30/72 (Q2)
<i>Studies in informatics and control</i>	National Institute for R&D in Informatics, ICI Bucharest	[11]	1.649	Operations research & management science (2020) = 61/84 (Q3); Automation & control systems (2020) = 43/63 (Q3)

Tabla 1.1: Revistas de los artículos publicados para la Tesis.

- Conferencia online titulada “*Mejorar la supervisión de la contratación pública: los algoritmos de IA aplicados a los datos en abierto de PLACSP*” impartida en la “*II Jornada del Observatorio Sector Público e IA de la Facultad de Derecho de la Univ. de Cádiz*” el 20/4/2022.
- Conferencia titulada “*Tecnologías digitales para velar por la contratación: el Big Data y la IA*” impartida en el “*XIII Seminario de Contratación Pública*”. Organizado por la Univ. de Zaragoza los días 22-23-24/6/2022 en Panticosa (Huesca).
- Conferencia titulada “*El análisis masivo de datos y la IA para mejorar el control y auditoría de la contratación pública*” impartida en el “*I Simposio Justicia y Derecho en Datos*”. Organizado por la Univ. de Salamanca y el Ministerio de Justicia los días 17-18/10/2022 en Salamanca.

- Conferencia titulada “*El análisis masivo de datos y la IA para mejorar el control y auditoría de la contratación pública*” impartida en el “X Congreso Nacional de Auditoría en el Sector Público”. Organizado por FIASEP los días 10-11/10/2022 en Tenerife.
- Conferencia titulada “*¿Es posible innovar en la contratación? Herramientas tecnológicas basadas en datos*” impartida en el “XII Congreso Internacional sobre contratación pública”. Organizado por la Univ. de Castilla-La Mancha los días 30-31/1/2023 en Cuenca.
- Conferencia titulada “*Cómo mejorar la contratación pública: digitalización, datos e inteligencia artificial*” impartida en las Jornadas “*La importancia de la contratación pública como motor de la economía*”. Organizado por la Diputación de Coruña el día 13/3/2023 en La Coruña.

CAPÍTULO SEGUNDO. RETOS Y OBJETIVOS

2.1. RETOS ACTUALES EN LA CONTRATACIÓN PÚBLICA

La contratación, como elemento principal del sector público para la prestación de servicios, no está exenta de los procesos de transformación y retos de la administración. Las AA.PP. han asimilado los procesos de transformación digital hasta el momento sin mayor inconveniente, aprovechando los instrumentos digitales para transformar la provisión de servicios y atención a la ciudadanía [17]. Sin embargo, todavía existen retos y áreas por cubrir, especialmente en el ámbito de la contratación.

Este estudio no trata de hacer un análisis exhaustivo de los retos que debe abordar la contratación pública, sino que identifica alguna de las problemáticas que se pueden abordar mediante la introducción de tecnologías. Es decir, utilizando las nuevas tecnologías como un medio o palanca de transformación para superar determinadas problemáticas, no como un fin en sí mismo. Podemos agrupar los retos en dos grandes bloques:

1. El aprovechamiento de los datos en la contratación para la mejora de las funciones de supervisión, monitorización, transparencia, planificación estratégica o la elaboración de políticas públicas.
2. La necesidad de realizar el proceso de compra pública de una forma más eficiente mediante la mejora operativa de la contratación.

En torno al primer bloque de retos, uno de los principales retos a los que se enfrenta la contratación es la calidad y la accesibilidad de los datos [18, 19]. En este sentido, la CE [20] destaca la falta de armonización de datos a nivel europeo como una de las grandes problemáticas para entender cómo están comprando y licitando las AA.PP. europeas. Esta falta de armonización se refleja en la imposibilidad de conocer, con precisión, cuánto están gastando las AA.PP. en la contratación de servicios, obras y suministros, o en la concesión de servicios públicos. Además, en España este problema se aprecia

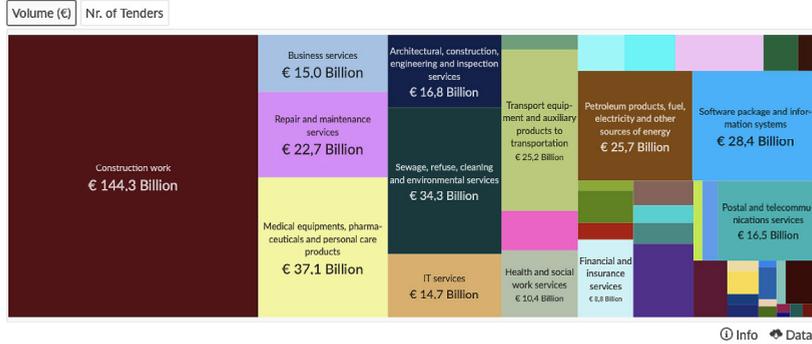
en la existencia de diversas fuentes de datos de contratación pública, con diferentes estándares de datos y de calidad y disponibilidad de la información. Esto se ha tratado de abordar en los últimos años mediante la obligatoriedad del uso de Plataforma de Contratación del Sector Público (PLACSP), pero todavía existe margen de mejora para garantizar el uso adecuado de la misma. En esta línea, el informe anual de la OIREscon sobre la contratación pública destaca el reto de la falta de información que permita la monitorización de la ejecución de contratos públicos [21], asignatura pendiente de las AA.PP. no sólo en España, sino en la mayoría de los estados europeos [22].

La supervisión y el control de la contratación es un pilar fundamental para garantizar el proceso de contratación, la buena ejecución de las compras públicas, el control de gasto público de manera eficaz y eficiente, elaboración de políticas públicas, así como prevenir y abordar problemas de corrupción y prácticas fraudulentas. En el sector privado se entiende que la contratación es una herramienta fundamental para conseguir los objetivos estratégicos de la organización y, sin embargo, esta idea apenas ha sido abordada en el sector público [4]. Para abordar este reto, diversas iniciativas se han implementado a nivel europeo, utilizando sistemas tecnológicos o innovadores [23]. En este sentido, la mayoría de los sistemas desarrollados hasta el momento han abarcado la detección de riesgos, asociado a la problemática de la corrupción. Se citan algunos ejemplos:

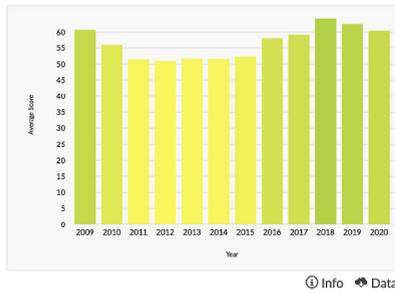
- El proyecto de [Open Tender](#) permite la búsqueda, análisis y supervisión de contratos públicos de los países Europeos. Ver un ejemplo para la supervisión del caso español en la Figura 2.1.
- El proyecto de [Red Flags](#), liderado por [Transparencia Internacional](#), permite la supervisión de contratos públicos en Hungría y la detección automática de alertas de riesgo.
- [DoZorro](#), liderado por varios actores internacionales, para la mejora del sistema de supervisión de la contratación pública de Ucrania.
- Iniciativas llevadas a cabo en Chile [24]. A través de su observatorio de compras públicas ha implementado diversas herramientas de medición de riesgo en las compras públicas de cara a realizar auditorías de contrato.

Otro reto es la transparencia en la gestión de recursos públicos ligada a la contratación. La transparencia es uno de los mecanismos fundamentales para garantizar las normas de funcionamiento de la UE con respecto al mercado único [25]. La introducción de reformas para conseguir transparencia en el gasto en contratación es crucial, facilitando el acceso a la información de precios, competencia o mecanismos de adjudicación [26]. Abordar este reto desde la adopción de reformas en el ámbito tecnológico es muy importante. Los beneficios que aportará la transparencia al proceso de contratación no están únicamente relacionados con la reducción de corrupción o prácticas ilícitas sino también mejorar la buena gobernanza de las administraciones públicas [27].

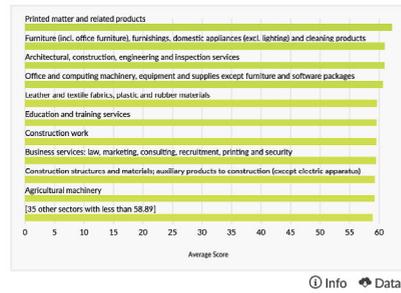
Sector Overview



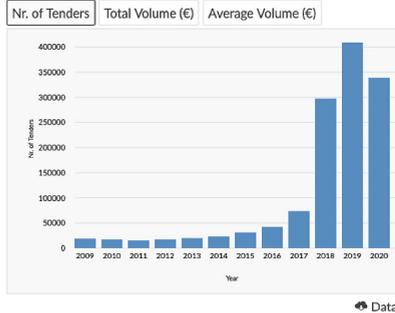
Average Good Procurement Score over Time



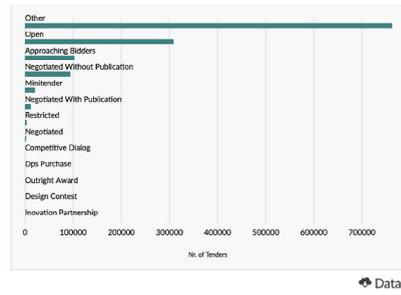
Average Good Procurement Score per Sector



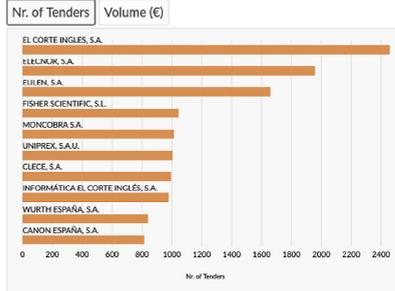
Tenders over Time



Procedure Type



Main Suppliers



Main Buyers



Figura 2.1: Ejemplo de supervisión de la contratación de España. Proyecto Open Tender.

En cuanto al segundo bloque de retos, el más relevante es la mejora operativa de los procesos de compra, teniendo en cuenta que el principal objetivo de la compra pública es la propia adquisición de bienes y servicios. Es un reto para las AA.PP. el tener información y un conocimiento extenso de temas tan relevantes para la consecución de una buena compra como son el diseño de la estrategia de contratación, la definición de los requisitos del servicio a prestar o el suministro a proveer, el conocimiento del propio mercado y los proveedores o la información sobre los precios de mercado [24]. La propia eficiencia operativa del proceso de contratación es algo perseguido por las AA.PP., de cara a ser más eficientes en la gestión de sus recursos. En este sentido, la automatización de procesos repetitivos o sin valor añadido o la reducción de la carga administrativa son materias pendientes de la contratación.

2.2. ORGANISMOS PÚBLICOS ESPAÑOLES RELACIONADOS CON CONTRATACIÓN

La innovación en las herramientas analíticas, basadas en datos como se proponen en esta Tesis, ayudarán a los diferentes organismos públicos relacionados con la gestión, fiscalización y supervisión de la contratación e, incluso, a la investigación policial. La ciudadanía cada vez es más consciente de la importancia de los contratos públicos, exigiendo mayores niveles en la gestión, transparencia y rendición de cuentas, sobre todo a los políticos que son la cabeza visible de la Administración. A continuación, se enumeran organismos públicos españoles que gestionan o utilizan información de contratos públicos:

- **Órganos y plataformas de contratación:**
 - Órganos de contratación en todos los niveles de la administración (local, autonómica y estatal) y las entidades dependientes del sector público. La contratación española está muy fragmentada, basta con dar el siguiente dato ilustrativo: en 2021 estaban registrados en Plataforma de Contratación del Sector Público (PLACSP) 20.500 órganos con competencias para realizar contratos.
 - PLACSP, plataformas de contratación autonómicas y la [Dirección General de Racionalización y Centralización de la Contratación \(DGRCC\)](#).
- **Organismos para la auditoría y control de la contratación:**
 - [Intervención General de la Administración del Estado \(IGAE\)](#) y las Intervenciones Generales autonómicas.
 - Oficinas Antifraude autonómicas.



Figura 2.2: Mapa de los organismos públicos españoles ligados a la contratación.

- Órgano de Control Externo Nacional (**Tribunal de Cuentas**) y autonómicos (Sindicaturas, Cámaras y Consejos de Cuentas).
- **Comisión Nacional de los Mercados y la Competencia (CNMC)**, así como Autoridades de la Competencia autonómicas.
- **Organismos que investigan judicialmente sobre contratación:**
 - Fuerzas y cuerpos de seguridad del Estado (**Guardia Civil**, **Policía Nacional**, etc.) que investigan a instancia de órganos judiciales.
 - **El Tribunal Administrativo Central de Recursos Contractuales** y los autonómicos.
- **Otros organismos con intereses en la contratación** y su transparencia son: **OIReScon**, **Junta Consultiva de Contratación del Estado** y las juntas autonómicas, **Autoridad Independiente de Responsabilidad Fiscal (AIReF)**, **Consejo de Transparencia y Buen Gobierno (CTBG)**, **Portal de la Transparencia de la AGE**, etc.

Por tanto, hay varias decenas de organismos (sin contar los miles de órganos de contratación) que se verían beneficiados si se dotasen de buenas herramientas analíticas, basadas en datos. Como en el mercado casi no existen programas informáticos para estos fines, las distintas AA.PP. tiene que diseñar y comprar aplicaciones informáticas ad hoc. Si se comprasen de manera conjunta y coordinada, se crearían muchas sinergias y se reducirían los costes de desarrollo, mantenimiento y evolución. Cuesta mucho esfuerzo diseñar y desarrollar aplicaciones informáticas pero luego son fácilmente adaptable a las distintos organismos y necesidades. Por ejemplo, las Comunidades Autónomas (CC. AA.) deberían de comprar conjuntamente aplicaciones informáticas específicas para sus órganos de contratación, Intervenciones Generales autonómicas, Órganos de Control Externo autonómicos, Comisiones de la Competencia autonómicas, etc. Como estos organismos tienen unas competencias autonómicas equivalentes, tendrán necesidades muy parecidas o iguales.

2.3. TRANSPARENCIA Y DATOS EN ABIERTO DE CONTRATACIÓN EN ESPAÑA

La transparencia en la contratación ha sido objeto de estudios jurídicos [28, 29, 30, 31, 26] y tiene tres aspectos fundamentales [32]. Primero, dar a conocer información relativa a los contratos públicos para rendir cuentas ante la sociedad. Segundo, contribuir a crear un mercado europeo de compras públicas. Y tercero, se obtendría información valiosa para incrementar la calidad de la gestión y supervisión, gracias a las técnicas de análisis masivo de datos. Además, una ventaja derivada es que tanto los órganos de contratación como los licitadores se sienten más vigilados, por lo que tienen más incentivos a hacer un trabajo más honesto y ajustado a derecho. En resumen, la transparencia actúa como principal “escudo” contra la corrupción [33]. Hay otros actores como los interventores, supervisores de la contratación, auditores de cuentas, autoridades de la competencia, etc. que también se benefician indirectamente de una mayor transparencia por convertirse en un ámbito de estudio mayor.

Para que haya transparencia se necesitan datos en abierto, llamados open data en su denominación inglesa. Son datos que pueden ser utilizados, compartidos y procesados libremente por cualquier persona, en cualquier lugar y para cualquier propósito (definición de la [Open Knowledge Foundation](#)). Los datos abiertos se basan en 8 principios: garantizar que están completos, primarios, oportunos, accesibles, procesables por una máquina, licencia libre, el acceso debe ser no discriminatorio y los formatos no deben ser propietarios. El estudio de temáticas asociadas a open data ha crecido muy fuertemente en el ámbito académico [34], sobre todo en la última década [35]. Esto es debido a la utilidad (directa o indirecta) de dichos análisis para obtener nueva información valiosa [36, 37] para los distintos involucrados. En el ámbito de la contratación pública, es cada vez mayor la cantidad y calidad de los repositorios de datos abiertos disponibles, tanto a nivel local, autonómico y estatal como europeo.

El open data asociado a la contratación pública está también evolucionando intensamente debido, principalmente, a factores tecnológicos (desarrollo de modelos y software para el e-Procurement [38]), burocráticos (estandarización del lenguaje de la contratación y digitalización de las AA.PP.), políticos (mayor transparencia en la toma de decisiones políticas), económicos (globalización, empresas compitiendo cada vez en mercados más lejanos a su origen) y sociales (menor tolerancia a la gestión pública y política, ineficaz y el interés por tener más información sobre las AA.PP. y sus contrataciones con empresas privadas). Para hacerse una idea de la importancia del open data en Europa, la Comisión Europea (CE) ha estimado [39] que en 2019 hubo un millón de trabajadores que trabajaron directamente con datos en abierto, el tamaño del mercado es de unos 180 mil millones de euros y las previsiones para los próximos años son de crecimientos significativos.

Un ejemplo ilustrativo del beneficio que aporta el open data es la licitación² que publicó la OIREscon en abril de 2021 cuyo objeto es el “*Desarrollo de herramientas ETL (Extract, Transform and Load) y modelo de datos de la información de múltiples plataformas de contratación*”. Para realizar sus estudios y trabajos de supervisión, la OIREscon contrata a una empresa para desarrollar un sistema informático que extraiga periódicamente los datos de las plataformas de contratación de España utilizando los datos en abierto. Dichos datos se deben validar, transformar y almacenar para que puedan ser explotados fácilmente por el personal de la OIREscon. Esta licitación indica la ausencia de mecanismos que hay para el intercambio de información entre las propias AA.PP., teniendo la OIREscon que recurrir a los datos en abierto. Y como la propia OIREscon indica en su último informe anual de 2021 [21]: “*se considera que debiera avanzarse a la generación de datos abiertos mucho más completos, en formatos homogéneos que faciliten su carga y tratamiento conjunto y respaldados por documentación acreditativa y técnica clara.*”

2.4. INICIATIVAS DE LA COMISIÓN EUROPEA ASOCIADAS A DATOS DE CONTRATACIÓN

En el primer apartado se mencionaba el reto de aprovechar los datos de contratación para mejorar las funciones de supervisión, monitorización, transparencia, planificación estratégica o la elaboración de políticas públicas. En la actualidad, la Comisión Europea (CE) tiene su portal de publicaciones de contratación, llamado Tenders Electronic Daily (TED), pero tienen carencias y limitaciones en relación a la publicación de licitaciones porque no contienen la misma información y, por tanto, no se asegura la homogeneidad y calidad de sus datos. Consecuentemente, la CE está trabajando en varias iniciativas

². El detalle de la licitación se puede consultar en: https://contrataciondelestado.es/wps/poc?uri=deeplink:detalle_licitacion&idEvl=Uwt%2FOsOe%2B5ymq21uxhbaVQ%3D%3D

para tener una verdadera estructura unificada de datos de contratación de los países miembros de la UE. El ecosistema digital promovido por la CE es un aspecto crucial para avanzar hacia un verdadero marco de contratación pública interoperable que garantice los principios del Mercado Único Digital (Single Digital Market) y, en particular, los del Portal Único Digital (Single Digital Gateway, SGD). Además, la alineación con eIDAS, habilitador clave para las transacciones electrónicas transfronterizas en la UE, debe considerarse como una oportunidad para facilitar la adopción de servicios digitales transfronterizos en el campo de la contratación pública.

Actualmente la CE está definiendo el espacio de datos de contratación pública (Public Procurement Data Space) (ver Figura 2.3) para que todos los países miembros de la UE vuelquen de manera estandarizada sus licitaciones y se pueda hacer un análisis de los datos a nivel europeo. Hasta ahora es muy difícil hacer análisis de datos de contratación a nivel europeo (en general, internacionalmente) porque cada país tiene su propia legislación y plataformas digitales de contratación. Por tanto, la CE es consciente del valor de los datos de contratación (al igual que el autor de esta Tesis) y para ello necesita previamente diseñar y crear un repositorio común de datos, alineando a todos los países miembros. Así, los distintos actores de la contratación podrán explotarlos convenientemente.

A continuación, se enumeran los sistemas o iniciativas más relevantes para conseguir el espacio de datos de contratación público de la UE:

- **Tenders Electronic Daily (TED)**. Es el portal de publicación de licitaciones de la UE.

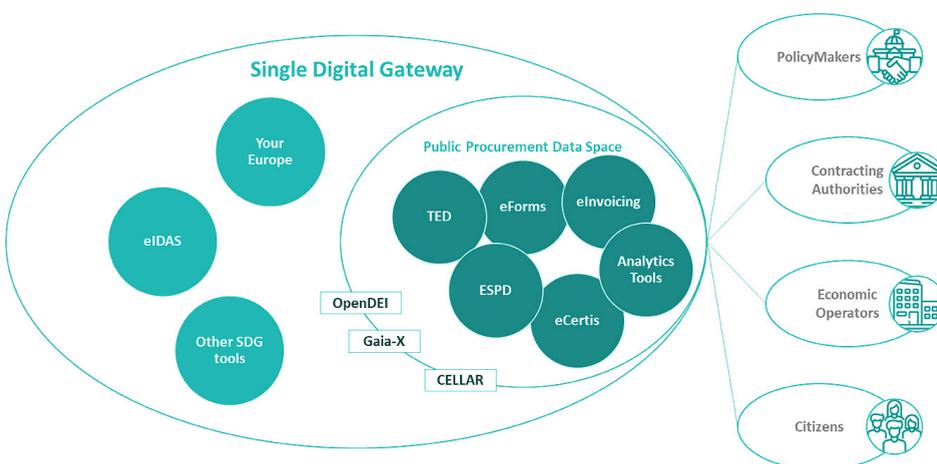


Figura 2.3: Elementos para tener un espacio de datos de contratación pública europeo unificado.

- **eForms**. Es una normativa³ de la CE para crear un estándar, formularios normalizados a rellenar por los órganos de contratación, en la publicación de las licitaciones en TED.
- **European Single Procurement Document (ESPD)**. Documento estandarizado de autodeclaración, según una normativa⁴, que es solicitado por los órganos de contratación a los operadores económicos para declarar que cumplen los requisitos legales y administrativos de la licitación en particular (si no los cumplen, serán excluidos del proceso de licitación).
- **eCertis**. Sistema de información que ayuda a los operadores económicos y órganos de contratación en la identificación y reconocimiento de los documentos y certificados más solicitados en los procedimientos de contratación de los Estados miembros. Persigue el objetivo de proporcionar interoperabilidad normativa y facilitar la contratación transfronteriza (órganos de contratación y operadores económicos de países diferentes).
- **eProcurement Ontology (ePO)**. Crear una ontología de contratación electrónica, comúnmente acordada, que codifique formalmente los procesos administrativos de contratación. Así se conseguirá un formato estructurado de datos de contratación legible por máquinas de manera automatizada. El objetivo es cubrir el proceso de contratación de principio a fin: publicación, proceso de licitación, adjudicación, pedido, facturación y pago. De esta manera, ePO conseguiría unificar las prácticas existentes de contratación electrónica, facilitando el intercambio, acceso y reutilización de datos.
- **CELLAR**. Es un repositorio semántico construido por la Oficina de Publicaciones de la CE. Es un servicio de infraestructura masivo que incluye grandes cantidades de contenido legal, publicaciones de la UE y conjuntos de datos de referencia. La plataforma se ha diseñado para ser escalable y admitir la interoperabilidad de los servicios, al proporcionar un marco común para el almacenamiento de contenido y metadatos.
- **eInvoicing**. Normativa europea⁵ para la facturación electrónica, debido a los distintos formatos de factura utilizados en los países miembros de la UE. Los órganos de contratación deberán aceptar las facturas que cumplan con la norma europea, pero seguirán siendo válidas las facturas que cumplan las diferentes normativas nacionales.

³. Reglamento de Ejecución (UE) 2019/1780 de la Comisión, de 23 de septiembre de 2019, por el que se establecen formularios normalizados para la publicación de anuncios en el ámbito de la contratación pública.

⁴. Reglamento de Ejecución (UE) 2016/7 de la Comisión, de 5 de enero de 2016, por el que se establece el formulario normalizado del documento europeo único de contratación.

⁵. Directiva 2014/55/UE del Parlamento Europeo y del Consejo, de 16 de abril de 2014, relativa a la facturación electrónica en la contratación pública

También hay iniciativas en el ámbito privado como, por ejemplo, [Gaia-X](#) que quiere crear una infraestructura de datos abierta, federada e interoperable, constituida sobre los valores de soberanía digital y disponibilidad de los datos. La CE tiene otra iniciativa llamada [OpenDEI](#), para la creación de plataformas de datos comunes basadas en una arquitectura unificada y un estándar establecido. En conclusión, se está en un momento de organización y estandarización de la información a nivel europeo, tanto en materia de contratación como en otros campos.

2.5. OBJETIVOS Y METODOLOGÍA DE LA INVESTIGACIÓN

En los siguientes 5 puntos se recogen los **principales objetivos de la Tesis**. Los 3 primeros se marcaron desde el inicio de la Tesis y los 2 últimos han sido consecuencia directa al ir desarrollándose los anteriores.

- **Adquisición y uso de los datos de contratación.** En 2017, cuando se comenzó la Tesis, era un reto en sí mismo el conseguir datos de licitaciones y poder manipular sus datos. En estos últimos 5 años en España se ha mejorado la cantidad, calidad y facilidad de acceso a los datos públicos de licitaciones. Por otro lado, esta Tesis hace hincapié en cruzar distintas fuentes de datos para abordar eficazmente los problemas en la contratación.
- **Empleo de innovadoras técnicas de Machine Learning (ML) a la contratación.** Al ser un campo donde se ha experimentado poco con técnicas de IA y ML, hay una incertidumbre al buscar problemas en los que se pueda aplicar con éxito estas técnicas.
- **Crear aplicaciones útiles en la contratación.** Es el objetivo más importante. Hay que desarrollar herramientas software que tengan interés para los agentes públicos y privados, obteniendo beneficios por su empleo, tanto para ellos como indirectamente para la sociedad. Se ha podido llevar a cabo combinando los dos puntos anteriores: utilizar fuentes de datos ligadas a la contratación y aplicar algoritmos de ML.
- **Detección de licitaciones irregulares.** Desarrollar programas informáticos que detecten licitaciones potencialmente fraudulentas para conseguir una contratación más segura y con menos corrupción. Y no sólo abordar el problema desde la visión técnica, sino también sacándolo a la luz pública para generar un debate mediático y señalar a los infractores. De esta manera se podrán corregir los problemas de fondo que afectan a la contratación en su conjunto.
- **Difusión de la investigación.** Al ser la contratación un tema de interés general para la sociedad y transversal para investigadores muy diversos, es recomendable hacer una labor de divulgación de la Tesis. Que la ciudadanía,

funcionarios, expertos, investigadores en contratación (juristas, economistas, politólogos, etc.) vean el potencial y utilidad del data science (ciencia de datos) y el ML. Para conseguir este fin, el autor ha impartido conferencias, dado entrevistas en medios de comunicación y presentado sus aplicaciones a premios (resultando premiado por el Gobierno Vasco, la Comisión Europea y el Tribunal de Cuentas de Galicia).

Todos los objetivos se han cumplido satisfactoriamente. En la actualidad se sigue con la labor de divulgación en foros especializados, con una cálida acogida por los expertos en contratación pública, principalmente juristas que son los que tradicionalmente han liderado esta disciplina.

A continuación, se resume la **metodología general** llevada a cabo en los artículos de la Tesis. Estas etapas son características del data science. Cada artículo ha requerido un desarrollo software ad hoc, realizado en el lenguaje de programación Python [3, 40, 41]. Para el detalle particular metodológico, léase el capítulo 4 donde se explica cada artículo.

1. **Formulación del problema y enfoque analítico.** Se sientan las bases, hipótesis, del objeto de estudio y se plantea el problema a resolver. En el enfoque analítico se identifica cuál sería el procedimiento que nos puede ayudar para obtener los resultados esperados.
2. **Adquisición de las fuentes de datos** (de contratación pública y otras fuentes asociadas). En general, se han utilizado datos públicos (PLACSP, TED) pero también datos no públicos (datos empresariales del Registro Mercantil y bases de datos de licitaciones colusivas).
3. **Almacenamiento** y estructuración de los datos. Se almacenan y estructuran los datos para poder ser analizados.
4. **Tratamiento de los datos.** Extracción, limpieza, transformación y filtrado de los datos. Enriquecimiento de los datos con fuentes complementarias cuando fuese posible.
5. **Comprensión de los datos.** Es decir, un análisis descriptivo y gráfico de los datos, no sólo para entenderlos sino también para conocer las carencias existentes. Se hace una verificación para asegurar la integridad y calidad de los datos que serán utilizados en la siguiente etapa. Esta etapa se itera, retroalimenta, a la etapa anterior.
6. **Desarrollo del modelo de predicción** (regresión o clasificación) utilizando algoritmos de ML. Dos de los artículos predicen el importe de adjudicación de la licitación. Otro artículo predice un grupo de empresas que pueden llevar a

cabo la licitación. Otro artículo clasifica la licitación en competitiva o colusiva (anticompetitiva, los ofertantes forman un cártel).

7. **Validación del modelo.** Se evalúan los algoritmos de ML mediante unas métricas de error (definidas en el capítulo 3) para seleccionar aquellos que más aciertan en la predicción.
8. **Discusión de los resultados y conclusiones.**

CAPÍTULO TERCERO. FUNDAMENTOS DE LA INVESTIGACIÓN

3.1. LA CONTRATACIÓN PÚBLICA

Actualmente la contratación pública en España es un área mucho más digital y cuantitativa que hace años. Es decir, los contratos tienen su información más relevante accesible digitalmente para su descarga masiva, de manera estructurada, pudiendo ser analizados dichos datos con herramientas propias de las disciplinas cuantitativas o científicas. Esta cuantización es debida a varios factores de origen legal:

- El primer paso básico es la digitalización de AA.PP. La Ley 39/2015 del Procedimiento Administrativo Común de las AA.PP. y la Ley 40/2015 de Régimen Jurídico del Sector Público vienen a configurar un escenario en el que la tramitación electrónica debe constituir la actuación habitual de las AA.PP.
- La LCSP unifica la publicidad y transparencia de los contratos públicos. También alude a que la publicación ha de realizarse en formatos abiertos y reutilizables, para que el ingente volumen de información pueda ser manejada por terceros.
- Transparencia y acceso a datos públicos por parte de los ciudadanos. La Ley 19/2013 de Transparencia, Acceso a la Información Pública y Buen Gobierno amplía y refuerza la transparencia de las AA.PP., regula y garantiza el derecho de acceso a la información relativa a sus actividades y establece las obligaciones de buen gobierno.
- La disposición de datos públicos crea un mercado que incentiva a empresas y personas a crear herramientas reutilizando dichos datos, obteniendo un beneficio económico por ello. Esto lo articula la Ley 18/2015 sobre la reutilización de la información del sector público.

Por tanto, estos cuatro factores legales posibilitan unas AA.PP. digitalizadas, publicidad y transparencia en la contratación, libre acceso a los datos de contratación y la posibilidad de reutilización de dichos datos con fines comerciales. La naturaleza descentralizada de la contratación en España, su gran dispersión y todavía el uso del papel en sus procesos no facilita el análisis digital y masivo de los datos. Sin embargo, si se echa la vista atrás en el tiempo el avance ha sido espectacular. Por ejemplo, en el año 2000 la manera de consultar las licitaciones era leyendo los Boletines Oficiales de las CC.AA. y del Estado. Hoy en día se pueden consultar de manera masiva y estructurada en internet a través de las diferentes plataformas de contratación autonómicas y nacional. El procesamiento masivo de datos se puede asociar al concepto tan manido de Big Data [42].

La contratación es ya un área cuantitativa como demuestra PLACSP (se explicará posteriormente en detalle), especialmente a partir de 2018. En 2021 se han publicado aproximadamente 200.000 licitaciones (considerando sólo perfiles del contratante en PLACSP y excluyendo contratos menores) y tiene datos de alta unos 20.500 órganos de contratación. Si se quiere conseguir una contratación moderna, eficiente y que esté bien controlada y supervisada, los organismos públicos que tienen asociadas estas responsabilidades deberían contar con equipos multidisciplinares. Desde juristas y economistas hasta ingenieros e informáticos que crean los sistemas (programas informáticos, aplicaciones web, bases de datos, etc.) que permiten analizar los datos de contratación de manera automática.

Las AA.PP. son cada vez más conscientes de la importancia del uso de la información para aumentar la eficiencia en sus procesos, ahorrar costes, aumentar la calidad de los servicios que prestan y para su toma de decisiones. La contratación genera un gran volumen de información, es decir, datos estructurados pero también documentos con información no estructurada (pliegos técnicos, administrativos, resoluciones de adjudicaciones, etc.). Es muy difícil de manejar ese volumen ingente de información por una persona u organismo que no tenga herramientas software especializadas.

3.2. LEGISLACIÓN EN LA CONTRATACIÓN PÚBLICA

Tanto en el ámbito europeo como en el español, se han desarrollado leyes relacionadas con la reutilización de la información del sector público y la contratación en el sector público que se resumen en la Tabla 3.1. Un tema particularmente importante en la contratación son las medidas anticorrupción, para así dificultar y frenar actuaciones fraudulentas, cuya normativa europea es objeto de la Tesis Doctoral de Javier Miranzo Díaz [43].

En el ámbito español, según la Ley 20/2013, disposición adicional tercera, y la LCSP, artículo 347, en la web de PLACSP se deben publicar todas las convocatorias de licitaciones y sus resultados por parte de todos los órganos de contratación que

Ley	Descripción	Ámbito	Enlace a la legislación
Directiva 2014/24/EU	Contratación pública	Europa	http://data.europa.eu/eli/dir/2014/24/oj
Directiva 2014/23/EU	Adjudicación de contratos de concesión	Europa	http://data.europa.eu/eli/dir/2014/23/oj
Directiva 2014/25/EU	Contratación por entidades que operan en los sectores del agua, la energía, los transportes y los servicios postales	Europa	http://data.europa.eu/eli/dir/2014/25/oj
Ley 9/2017	Transposición al Derecho español de las anteriores Directivas europeas 2014/23/UE y 2014/24/UE	España	https://boe.es/eli/es/l/2017/11/08/9
Directiva 2003/98/EC	Reutilización de la información del sector público	Europa	http://data.europa.eu/eli/dir/2003/98/oj
Directiva 2013/37/EU	Modificación de la anterior Directiva 2003/98/CE	Europa	http://data.europa.eu/eli/dir/2013/37/oj
Directiva 2007/2/EC	Establecimiento de una Infraestructura de Información Espacial en la Comunidad Europea (INSPIRE)	Europa	http://data.europa.eu/eli/dir/2007/2/oj
Ley 37/2007	Transposición al Derecho español de la Directiva Europea 2003/98/CE	España	https://boe.es/eli/es/l/2007/11/16/37
Real Decreto 1495/2011	Desarrollo de la Ley española 37/2007	España	https://boe.es/eli/es/rd/2011/10/24/1495
Decisión de la CE 2011/833/EU	Sobre la reutilización de los documentos de la CE	Europa	http://data.europa.eu/eli/dec/2011/833
Ley 19/2013	Transparencia, acceso a la información del sector público y buen gobierno	España	https://boe.es/eli/es/l/2013/12/09/19
Ley 20/2013	Garantía de la unidad de mercado	España	https://boe.es/eli/es/l/2013/12/09/20
Ley 18/2015	Transposición al Derecho español de la Directiva Europea 2013/37/UE	España	https://boe.es/eli/es/l/2015/07/09/18

Tabla 3.1: Legislación sobre contratación pública y el uso de datos.

pertenecen al Sector Público estatal. Según la LCSP, las CC.AA. podrán optar por publicar sus perfiles del contratante a través de sus propios servicios de información o directamente en PLACSP. Si optaran por publicar su información contractual en medios ajenos a PLACSP, deberán publicar en la misma mediante mecanismos de agregación las publicaciones de las licitaciones y sus resoluciones. Las Administraciones locales, así como sus entidades vinculadas o dependientes, podrán optar por alojar la publicación de sus perfiles de contratante en el servicio de información que a tal efecto estableciera la Comunidad Autónoma de su ámbito territorial, o bien por alojarlos en PLACSP. Si optaran por alojarlos en el servicio de la Comunidad Autónoma, esta deberá publicar mediante mecanismos de agregación las convocatorias de licitaciones y sus resultados en PLACSP.

La información mínima que deben contener los anuncios de las licitaciones está definida en la LCSP, Anexo III “*Información que debe figurar en los anuncios*”. PLACSP dispone de un apartado de Open Data para la reutilización de información de las licitaciones publicadas, en cumplimiento de las obligaciones de publicidad establecidas en la LCSP.

Respecto al anuncio oficial de las licitaciones que requieren publicidad fuera del ámbito español, el artículo 135 de la LCSP establece que cuando las licitaciones estén sujetas a regulación armonizada (es decir, aquellas licitaciones con un importe mayor a un determinado umbral o con ciertas características, estipuladas en los artículos del 19 al 23 de la LCSP), la licitación deberá publicarse, además, en el [Diario Oficial de la Unión Europea \(DOUE\)](#). Cuando el órgano de contratación lo estime conveniente, se podrán anunciar en el DOUE las licitaciones no sujetas a regulación armonizada.

La UE tiene un [portal oficial de datos abiertos](#) que se creó en 2012, de conformidad con la Decisión de la CE 2011/833/EU sobre la reutilización de documentos de la CE, e invita a todas las instituciones europeas a que pongan sus datos a disposición del público siempre que sea posible. Además, existe un portal denominado llamado TED dedicado a la contratación pública europea. En este portal, se publican diariamente todos los anuncios de licitación y adjudicación, además de anuncios de información previa, para contratos sujetos a regulación armonizada (S.A.R.A.), es decir, que son susceptibles de ser de interés transfronterizo y para los cuales se debe garantizar la libre concurrencia de operadores de todos los estados miembro de la UE. En cuanto a la reutilización de los datos de TED de cara a producir informes sobre el estado de la contratación pública en Europa, es la Dirección General de Comercio (DG GROW) la que se encarga de la elaboración de un informe de indicadores anuales sobre contratación pública⁶ y el mantenimiento de un portal de indicadores comparativos a nivel europeo⁷.

⁶ La última versión de los indicadores aporta información sobre la contratación del año 2017: <https://ec.europa.eu/docsroom/documents/38003>

⁷ Accesible en https://ec.europa.eu/internal_market/scoreboard/performance_per_policy_area/public_procurement/index_en.htm

Además, se realizan estudios ad-hoc⁸ sobre el estado del arte de la contratación que tienen como objetivo influir en la elaboración de políticas públicas, proporcionando un mayor conocimiento sobre el mercado público y su funcionamiento.

Queda fuera del alcance de la Tesis explicar los principios generales del derecho de la contratación pública internacional y la comparación de las legislaciones nacionales (el llamado Derecho comparado). Esta legislación define el sistema de contratación de cada país y, por tanto, debería articular los mecanismos de publicación de los datos de los contratos públicos (transparencia). Para ahondar en esta complejo tema jurídico se puede consultar [44].

Un debate que está en boga en el ámbito jurídico es la ética de los algoritmos de IA en su funcionamiento, produciendo sesgos o búsquedas dirigidas [45]. La realidad es que los algoritmos aplicados a la contratación están en un estado incipiente, poco evolucionado. Por lo tanto, la ética de la IA en este campo está de momento en un plano teórico, filosófico, hasta que dichos algoritmos no se sofisticuen y tengan un comportamiento más complejo, más humano.

3.3. TECNOLOGÍAS EN LA CONTRATACIÓN PÚBLICA

A continuación, se enumeran una serie de relevantes tecnologías software de reciente desarrollo o difusión en el ámbito de las AA.PP. Son la base tecnológica para aumentar la eficiencia de la contratación, facilitando nuevas herramientas que permitan dar un salto cuantitativo y cualitativo. Se han agrupado en 3 bloques: data, IA y automatización de procesos:

- **Data**

- **Procesamiento masivo de datos (Big Data)**. Se ocupa de las actividades relacionadas que manipulan grandes conjuntos de datos o que requieren gran capacidad de computación: extracción masiva de datos, almacenamiento y búsqueda, software optimizado para trabajar en paralelo con gran cantidad de información, computación en la nube, etc...El límite de procesamiento ha ido creciendo a lo largo de los años, no es un concepto estático. Los ordenadores de consumo actuales son supercomputadores si se comparan con los de hace 25 años. Los sistemas informáticos de contratación no son especialmente intensivos en el almacenamiento o intercambio de datos como pueden ser otros (banca, compañías telefónicas, empresas de contenidos audiovisuales bajo demanda, etc.). Por tanto, las tecnologías software llamadas Big Data no son estrictamente necesarias en contratación por el volumen de información manejada, de momento, pero sí recomendables para tener un sistema escalable que funcione con una mayor cantidad de datos.

⁸. Accesibles en https://ec.europa.eu/growth/single-market/public-procurement/studies-networks_en

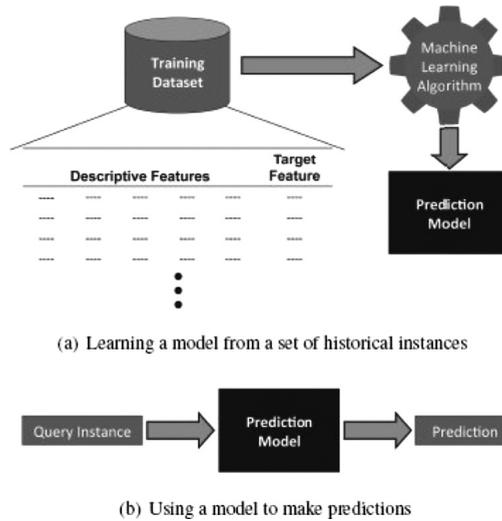


Figura 3.1: Funcionamiento de un algoritmo de ML: entrenamiento (a) y uso (b).
Fuente [46].

– **Análisis de datos (Data Analytics)**. Es un proceso que consiste en inspeccionar, limpiar y transformar datos con el objetivo de que un usuario pueda visualizar y comprender la información para detectar o resaltar información útil, dando lugar a conclusiones que apoyen la toma de decisiones. Ejemplos comerciales de este tipo de herramientas software son Microsoft Power BI, Tableau o Kibana. En nuestro caso, se pueden utilizar para transformar los datos de contratación y crear cuadros de mando con indicadores y gráficas para ayudar en la operación y supervisión de la contratación.

- **Inteligencia Artificial (IA):**

– **Machine Learning (ML)**. Son algoritmos informáticos para el descubrimiento de nuevo conocimiento a partir de grandes cantidades de datos. Es decir, es un proceso automático para extraer patrones de los datos [46] (ver Figura 3.1). El “aprendizaje” entra en juego cuando damos a estos algoritmos parámetros adaptables a los datos observados (el programa está “aprendiendo” de los datos). Una vez que estos modelos se han entrenado, ajustado, a esos datos, se pueden usar para predecir y comprender aspectos de otros datos nuevos [3]. Se considera una de las innovaciones más disruptivas y un fuerte factor para crear ventajas competitivas. Si bien el ML ha existido durante más de 60 años, recientemente ha mostrado un potencial significativo para cambiar las economías y sociedades [47]. Es la principal tecnología usada en esta Tesis.

– **Procesamiento del Lenguaje Natural (NLP)**. Se ocupa de la formulación e investigación de mecanismos eficaces computacionalmente para la comunicación entre personas y máquinas por medio del lenguaje natural. En la contratación pública una gran cantidad de información (con mucho valor) se almacena en documentos no estructurados (pliegos, actas, memorias, etc.), en lenguaje natural. Esta tecnología es especialmente relevante para obtener información estructurada/clasificada a partir de textos, comprensión del lenguaje natural por parte de la máquina (NLU) para poder hacer búsquedas por tópicos y no por palabras exactas, respuestas elaboradas por una máquina a preguntas de una persona (chatbot), traducción automática (textos en idiomas distintos), etc.

- **Automatización de Procesos:**

– **Minería de Procesos (Process mining)**. Análisis de los procesos de la información utilizando el registro de eventos entre las distintas etapas del flujo de información. Esta tecnología permite analizar la traza de los procesos en estudio, incluyendo información de los actores que lo realizan, los tiempos involucrados, volumen que pasa por cada etapa, cuellos de botella, los caminos típicos que se recorren, etc. Uno de los objetivos es llevar el control de los procesos, pero además tiene como objetivo permitir el descubrimiento y optimización de procesos, controles, información y estructuras organizacionales partiendo de dichos registros de eventos. Los sistemas informáticos de contratación pública son buenos destinatarios de esta tecnología, les ayudaría a mejorar sus procesos de licitación e, incluso, etapas posteriores: seguimiento del contrato, supervisión y facturación.

– **Automatización Robótica de Procesos (RPA)**. Es la automatización de los procesos/etapas de negocio que realiza una máquina replicando las acciones que hace una persona interactuando con la interfaz de usuario (IU) del sistema informático. Es decir, el robot software opera en la IU de la misma manera que un ser humano. Esto es una diferencia significativa a las formas tradicionales que se basan en Interfaces de Programación de Aplicación (API). El RPA es una tecnología que aplicada convenientemente permitiría aumentar la eficiencia de los órganos de contratación, automatizando validaciones, cálculos o tareas repetitivas en las distintas etapas del proceso de contratación.

3.4. APLICACIONES EN LA CONTRATACIÓN PÚBLICA

En este apartado se explican una serie de casos de uso relevantes para la contratación utilizando las nuevas tecnologías anteriormente descritas. Se ha dividido en 3 bloques (ver Figura 3.2): fuentes de datos, procesamiento y almacenamiento y posibles aplicaciones.

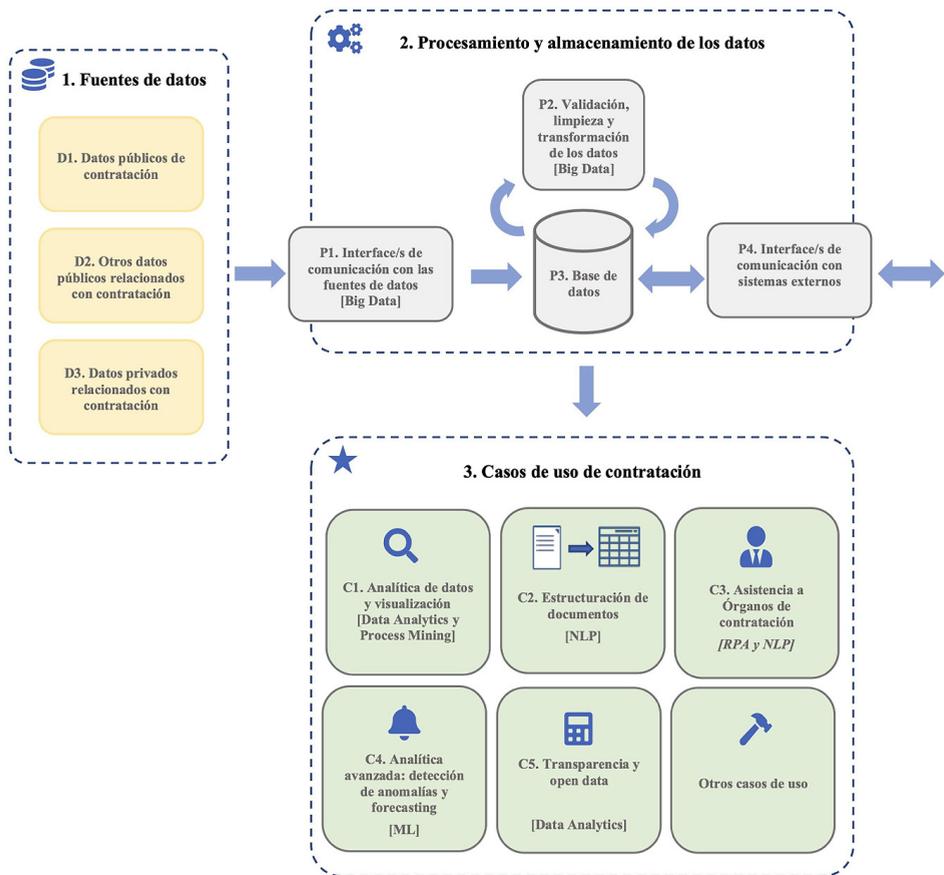


Figura 3.2: Diagrama de posibles aplicaciones para la contratación pública.

1. **Fuentes de datos.** El primer paso es disponer de un gran repositorio de información de contratación que permita tener una visión completa. Para ello se identifican diferentes fuentes de datos, tanto públicas como privadas, que pueden contribuir a generar dicho repositorio. Se pueden dividir en tres grupos según su origen:

- **D1. Datos públicos de contratación.** Principalmente, en España serían las plataformas de contratación de las CC.AA., la nacional PLACSP y la DGRCC (gestiona acuerdos marco, sistemas dinámicos de adquisición y contratos centralizados). A nivel europeo, la plataforma de contratación de los países miembros es TED.
- **D2. Otros datos públicos relacionados con contratación.** Por ejemplo, resoluciones de la Junta Consultiva de Contratación Pública del Estado, el

Tribunal Administrativo Central de Recursos Contractuales, la [Central de Información Económico-Financiera de las AA.PP.](#), el Portal de Transparencia de la AGE, el [Registro de Contratos](#) o la [Base Nacional de Subvenciones](#) para rastrear las ayudas concedidas a personas jurídicas o físicas.

- **D3. Datos privados relacionados con contratación.** Por ej., información operacional de acceso privado de las plataformas de contratación antes mencionadas o los datos almacenados en el [Registro Mercantil](#) para obtener información empresarial y económica de los adjudicatarios.

2. Procesamiento de los datos de contratación. Los componentes fundamentales serían:

- **P1. Interface de comunicación con las fuentes de datos.** La función principal es la automatización de la ingesta de los datos. Para las fuentes de tipo D1 serían los datos estructurados de la licitación y los documentos de la licitación (pliegos técnicos y administrativos, anexos, etc.).
- **P2. Validación, limpieza y transformación de los datos.** Hay que validar la coherencia e integridad de los datos, limpiarlos y enriquecerlos gracias a las distintas fuentes. En este proceso se harían todas las tareas necesarias para normalizar y transformar los datos.
- **P3. Base de datos.** Almacena la información recibida de las distintas fuentes (datos brutos) y, tras procesarlos en el componente P2, almacena también los datos netos (transformados, enriquecidos) que se utilizarán en los diferentes casos de uso. De esta manera, se unifica toda la información en un único repositorio. En cuanto a la tecnología, puede ser desde una base de datos de tipo SQL tradicional a específicas de Big Data si el volumen de datos lo requiere.
- **P4. Interface de comunicación con sistemas externos.** En el futuro la intercomunicación de los sistemas será mucho mayor que en la actualidad y se crearán mayores sinergias por la interconexión entre los diferentes organismos con competencias en contratación.

3. Aplicaciones. Gracias al procesamiento y almacenamiento de los datos realizado en la fase anterior, ya se tiene la infraestructura necesaria para llevar a cabo los casos de uso. A continuación, se ponen varios ejemplos aunque el abanico de casos es muy grande (más ejemplos en [48]), tantos como necesidades tengan los involucrados de la contratación.

- **C1. Analítica de datos y visualización.** Visualización de información analítica de contratación (tablas, gráficos, indicadores, informes): análisis geográfico de los lugares de ejecución de las licitaciones, sectores con mayor contratación,

- plazos en la resolución de la adjudicación, competencia empresarial (número de empresas que participan en una licitación), características de las empresas que más licitaciones ganan (tamaño, origen, sector principal de negocio, tipo de sociedad mercantil, etc.), precios máximos y mínimos de las ofertas, desviación del importe de licitación respecto al importe de adjudicación (ahorro obtenido), cómo y qué contratan los órganos contratantes, impacto de la división en lotes de los contratos (y la fragmentación ilícita en contratos menores), comparación de parámetros de contratación con años pasados y con otras Administraciones, etc.
- **C2. Estructuración de documentos.** El proceso de contratación en gran parte se basa en la gestión documental. Esto supone que la información relevante se encuentra en documentos (pliegos, formularios, anexos, etc.) de manera desestructurada. Los documentos deben ser correctamente interpretados, clasificados y estructurados para almacenarse en la base de datos y ser explotados posteriormente. De esta manera se tiene una visión completa y se podrán crear aplicaciones que ayuden en el análisis y toma de decisión.
 - **C3. Asistencia a órganos de contratación.** Los órganos de contratación suelen trabajar de manera independiente sin mucha interrelación, realizando consultas a otros departamentos de asistencia en materia de contratación. Es decir, en las AA.PP. suele haber una fragmentación de la contratación coexistiendo muchos órganos que se enfrentan licitaciones que nunca antes había redactado ni tienen herramientas comunes para trabajar colaborativamente. Ejemplos de aplicaciones que les ayuden serían la automatización de las etapas en el proceso de contratación (RPA), chatbot para dar respuestas a preguntas recurrentes, herramientas colaborativas para reaprovechar el conocimiento de los órganos, generación automatizada de un repositorio común de proveedores u otra información de interés, etc.
 - **C4. Analítica avanzada: detección de anomalías y forecasting.** El control de la contratación se hace generalmente a partir de la revisión de expedientes de contratos individuales, lo que dificulta la detección de eventuales prácticas fraudulentas. No se pueden hacer estimaciones a futuro por no tener herramienta que analicen de manera agregada las licitaciones. Sin embargo, un tratamiento masivo de la información disponible de contratación permite hacer forecasting, identificar patrones y detectar anomalías que ayuden a identificar prácticas irregulares. Los artículos de investigación realizados en esta Tesis son precisamente de este tipo, son buenos ejemplos del uso de algoritmos de IA y ML.



Figura 3.3: Web de la Plataforma de Contratación del Sector Público (PLACSP).

- **C5. Transparencia y open data.** Hay pocos portales de transparencia donde se muestre información detallada de contratación (tanto a nivel agregado como desagregado) a los ciudadanos como ejercicio de transparencia y rendición de cuentas. Para resolver esto, habría que simplificar el primer caso de uso (C1), mostrando datos y visualizaciones generales que se pueden descargar libremente para que ciudadanos o usuarios externos a las AA.PP. puedan hacer sus propios análisis e investigaciones.

3.5. LA PLATAFORMA DE CONTRATACIÓN DEL SECTOR PÚBLICO (PLACSP)

El matemático Clive Humby afirmó que “*Los datos son el nuevo petróleo. Los datos son valiosos pero tienen que refinarse como el petróleo, si no carecen de utilidad*”. La contratación pública genera un gran volumen de información, datos en crudo que necesitan refinarse para aportar utilidad. La mayor plataforma de datos de contratación en España es PLACSP (ver Figura 3.3), así como en la UE es el TED. En las siguientes subsecciones se detallará PLACSP desde una perspectiva del origen de los datos, su formato y su uso.

3.5.1. Órganos de contratación que publican en PLACSP

La PLACSP, regulada en el artículo 347 de la ley LCSP, es: “*...una plataforma electrónica que permita la difusión a través de Internet de sus perfiles de contratante, así como prestar otros servicios complementarios asociados al tratamiento informático de estos datos*”. Este mismo artículo define a la PLACSP como un instrumento de publicidad obligatorio para todo el Sector Público Estatal. Define la posibilidad que tienen las CC.AA. y las Ciudades Autónomas para establecer servicios de información similares pero que deberán también publicar las convocatorias de licitaciones y sus resultados en PLACSP. Finalmente, en el último párrafo del punto 3 del artículo 347, se indica dónde deben estar alojados los perfiles de contratante de las administraciones locales y sus entidades vinculadas o dependientes.

En resumen, se establecen tres ámbitos con distintos niveles de obligatoriedad según la legislación vigente de la LCSP:

- **Sector Público Estatal.** Todos los órganos de contratación están obligados a publicar su perfil del contratante en PLACSP, lo que incluye todos los documentos e informaciones referentes a su actividad contractual, y en concreto: anuncios de licitación, pliegos, anuncios de adjudicación, anuncios de formalización, anuncios de renuncia o desistimiento u otros documentos.
- **Comunidades Autónomas (CC.AA.) y Ciudades Autónomas.** Todos los órganos de contratación, así como sus entes, organismos y entidades vinculadas deberán alojar sus perfiles de contratante en el sistema de información que disponga la CC.AA. o Ciudad Autónoma y comunicar a PLACSP mediante el mecanismo de agregación un conjunto reducido de datos sobre la publicación de la convocatoria y su resolución. En caso de no disponer de dicho sistema de información, los perfiles de contratante deberán alojarse en PLACSP.
- **Administraciones locales.** Pueden optar por adherirse a PLACSP o a la que disponga su CC.AA. o Ciudad Autónoma.

Las licitaciones publicadas en PLACSP aumentaron significativamente a partir de 2018 debido a la entrada en vigor de la LCSP en 2017. Progresivamente se han ido incorporando más órganos de contratación a dicha plataforma.

3.5.2. Origen de los datos de PLACSP

La PLACSP tiene varios conjuntos de datos abiertos referentes a las licitaciones publicadas en el portal de transparencia del Ministerio de Hacienda. Todos estos conjuntos de datos abiertos son accesibles⁹ y diariamente se publican las actualizaciones del día anterior. Concretamente, se publican 3 conjuntos de datos abiertos (ficheros XML con extensión .atom¹⁰):

1. **Sector Público.** Expedientes de contratación publicados en los perfiles del contratante ubicados en PLACSP, excluyendo los contratos menores.
2. **Agregadas.** Licitaciones publicadas en PLACSP mediante mecanismos de agregación, excluyendo los contratos menores.
3. **Menores del Sector Público.** Contratos menores publicados en los perfiles del contratante ubicados en PLACSP.

⁹. Accesible en https://www.hacienda.gob.es/es-ES/GobiernoAbierto/Datos%20Abiertos/Paginas/licitaciones_plataforma_contratacion.aspx

¹⁰. La estructura del fichero XML se conforma siguiendo las especificaciones descritas para un “Atom Feed Document” en la RFC 4287 de la IETF. Accesible en <https://www.ietf.org/rfc/rfc4287.html>

Las administraciones autonómicas que opten por mantener los perfiles del contratante en su propia plataforma de contratación deberán publicar las convocatorias de licitación y sus resultados en PLACSP mediante mecanismos de agregación. En la actualidad, las plataformas de contratación que van al conjunto de datos 2 (agregadas) son las 7 siguientes CC.AA.: Cataluña, País Vasco, La Rioja, Comunidad de Madrid, Galicia, Andalucía y Navarra. El resto de CC.AA., incluyendo a Ceuta y Melilla, están alojadas en PLACSP. Es decir, van al conjunto de datos 1 (Sector Público) o 3 (contratos menores del Sector Público). Hay más organismos públicos (Administración General del Estado, Mutuas, Entidades Locales, Universidades, etc.) que también están integrados en sendos conjuntos de datos. El artículo 63 de la LCSP en su apartado cuarto establece que la información del perfil de contratante se publicará en formatos abiertos y reutilizables, permaneciendo accesible al público durante un tiempo no inferior a 5 años.

3.5.3. Formato y calidad de los datos en abierto de PLACSP

Todo el detalle referente al formato de los ficheros de datos abiertos y a los campos disponibles en las licitaciones se puede consultar en el “*Formato de sindicación y reutilización de datos sobre licitaciones publicadas en la Plataforma de Contratación del Sector Público*”, siendo su última versión¹¹ de fecha 19/10/2021. Dicho documento lo elabora la Subdirección General de Coordinación de la Contratación Electrónica (D. G. del Patrimonio del Estado). Para más detalle, consultar el documento “*Resumen de contenido en conjuntos de datos abiertos*” donde especifica más de 100 campos disponibles¹² que se agrupan dentro de las siguientes áreas: datos generales del expediente, lugar de ejecución, lotes, procesos de licitación, entidad adjudicadora, plazo de obtención de pliegos, extensión del contrato, condiciones de licitación, garantías requeridas, requisitos de participación, criterios de evaluación técnica y económica-financiera, subcontratación permitida, criterios de adjudicación, limitación del número de licitadores a invitar, información sobre el contrato, adjudicatario, importe de adjudicación, condiciones de subcontratación, justificación del proceso, modificaciones del contrato, publicaciones oficiales y otros documentos publicados.

La información sobre los expedientes de licitación se van agregando de forma incremental utilizando la arquitectura Componentes y Documentos Interoperables para la Contratación Electrónica (CODICE)¹³. Ésta proporciona una biblioteca de componentes estándar, reutilizables, y extensibles o adaptables a diversos contextos o necesidades de contratos públicos específicos, para satisfacer las necesidades de información de los documentos y mensajes intercambiados a lo largo del ciclo completo de los procedimientos electrónicos de contratación. Las especificaciones

¹¹ Accesible en <https://contrataciondelsectorpublico.gob.es/datosabiertos/especificacion-sindicacion.pdf>

¹² Accesible en https://contrataciondelsectorpublico.gob.es/datosabiertos/DGPE_PLACSP_ResumenDatosAbiertos.pdf

¹³ Más información de CODICE en <https://contrataciondelestado.es/wps/portal/codice>

CODICE están incluidas en el catálogo de estándares del Esquema Nacional de Interoperabilidad.

Tras manejar los datos de PLACSP durante los últimos 5 años para poder elaborar esta Tesis, se ha llegado a la conclusión que PLACSP no tiene mecanismos para validar el formato y coherencia de los datos que se introducen en cada licitación. La plataforma es un contenedor de información, siendo los órganos de contratación los únicos responsables legales en la veracidad y completitud de la información de las licitaciones que suben a la plataforma. Este extremo se menciona por la OIREscon [21] y ha sido confirmado por los funcionarios responsables de PLACSP¹⁴. No se ha encontrado legislación que castigue o penalice a los órganos de contratación que cometan errores en los datos publicados de sus licitaciones. Este hecho será muy dañino para los investigadores que traten estos datos. Para más detalle sobre la transparencia y los datos en abierto de la PLACSP y las plataformas autonómicas de contratación, consultar [21].

Como se acaba de decir, que un campo exista (mencionados anteriormente) no significa que se haya rellenado, que tenga el formato correcto o que el valor se haya introducido correctamente. Todos estos casos se han encontrado en PLACSP, especialmente campos no rellenados, y consecuentemente ha dificultado la realización de los artículos de esta Tesis. Ciertos campos deben tener una estructura fija (NIF, código postal, fecha, etc.) pero algunos valores no siguen dicha estructura. Además, se han detectado valores anómalos. Por ejemplo, que el número de ofertantes tenga un valor altísimo (más de 1.000 ofertantes, suceso extremadamente improbable) o que el importe de adjudicación sea muy superior al de licitación (50 veces superior al de licitación). Consecuentemente, la calidad de los datos en PLACSP es baja, se deben de “limpiar” (data cleaning explicado en [10]) para conseguir estudios rigurosos, sin información errónea que desvirtúe el análisis de datos agregados.

Hay que dotar a PLACSP de mecanismos de validación automático cuando el funcionario del órgano de contratación introduzca los datos de la licitación. Es comprensible los errores humanos y muchos de ellos se podrían corregir de manera sencilla. Por otro lado, se deberían de incluir en las leyes de contratación (LCSP y afines) una regulación sobre la calidad del dato y las responsabilidades legales que se deriven si los órganos de contratación las incumplen.

La falta de calidad del dato, interoperabilidad, poca estandarización, etc. no es exclusivo de PLACSP o del ámbito de la contratación, son carencias generalizadas en las AA.PP. españolas. Para paliar esto se crea la Oficina del Dato¹⁵ en 2020,

¹⁴. María Lafuente Fernández, Subdirectora General Adjunta de Coordinación de la Contratación Electrónica y encuadrada en la D.G. de Patrimonio del Estado (Ministerio de Hacienda y Función Pública), como responsable de PLACSP mencionó esta cuestión en el encuentro “*Contratación Electrónica Inteligente: situación y retos*” en el Instituto Nacional de Administración pública celebrado el día 20/1/2022.

¹⁵ Orden ETD/803/2020 publicada en el BOE: https://www.boe.es/diario_boe/txt.php?id=BOE-A-2020-10008

dependiente de la Secretaría de Estado de Digitalización e IA, siendo sus principales líneas estratégicas “*el diseño, coordinación y seguimiento del modelo de referencia arquitectónico para fomentar la recolecta, gestión e intercambio de datos públicos*”. España necesita progresar mucho en esta materia y esta oficina jugará un papel capital si se le dota del personal y recursos adecuados. Hay que tomar referentes como, por ejemplo, el Reino Unido que publicó una guía gubernamental de calidad del dato¹⁶ en 2020.

3.5.4. Programa OpenPLACSP para analizar los datos en abierto

La Subdirección General de Coordinación de la Contratación Electrónica ha puesto a disposición del público en el verano de 2021 el programa OpenPLACSP¹⁷ (ver captura de pantalla en la Figura 3.4) para facilitar la transformación de los ficheros de datos abiertos en una tabla (hoja de cálculo). Con esta herramienta se pretende que cualquier interesado pueda trabajar de una forma rápida y sencilla con las licitaciones publicadas. Esto no era así hasta la aparición de este programa. Se debían de tener conocimientos de programación para convertir los ficheros XML (.atom) a una base de datos u hoja de cálculo y poder explotar los datos. Como los artículos de esta Tesis que utilizan datos de PLACSP son de fecha anterior al verano de 2021, se tuvo que desarrollar un script ad-hoc para convertir dichos ficheros .atom.

Este programa es un gran avance para organizaciones, investigadores o ciudadanos que quieran realizar análisis y estudios de contratación sin tener conocimientos de programación. Es un elemento necesario para que cada vez haya más estudios jurídico-económico de tipo cuantitativo. ¿Qué tipo de procedimiento produce mayores bajas económicas? ¿Y cuántas ofertas se tienen de media? ¿Favorece la competencia o no? Se podrían hacer un sinnúmero de preguntas.

3.6. LITERATURA ACADÉMICA SOBRE LA CONTRATACIÓN PÚBLICA

Las tecnologías como el ML y el big data se pueden (y deben) aplicar al campo económico por cosechar éxitos notables [49, 50, 51]. Esta Tesis particulariza su aplicación al ámbito de la contratación, campo muy relevante por su peso en el PIB y por ser un gasto público, que concierne a toda la sociedad. Hay abundante bibliografía para conocer en detalle los algoritmos de ML y la matemática que hay detrás de ellos [52, 53]. Sin embargo, no hay tantas referencias sobre su aplicación en la contratación por ser un campo todavía poco explorado e investigado desde el ámbito científico-tecnológico.

¹⁶. “The Government Data Quality Framework: principles, guidance and case studies”. Accesible en <https://www.gov.uk/government/publications/the-government-data-quality-framework>

¹⁷. Accesible para su descarga en la web de PLACSP en el menú “Datos abiertos”.

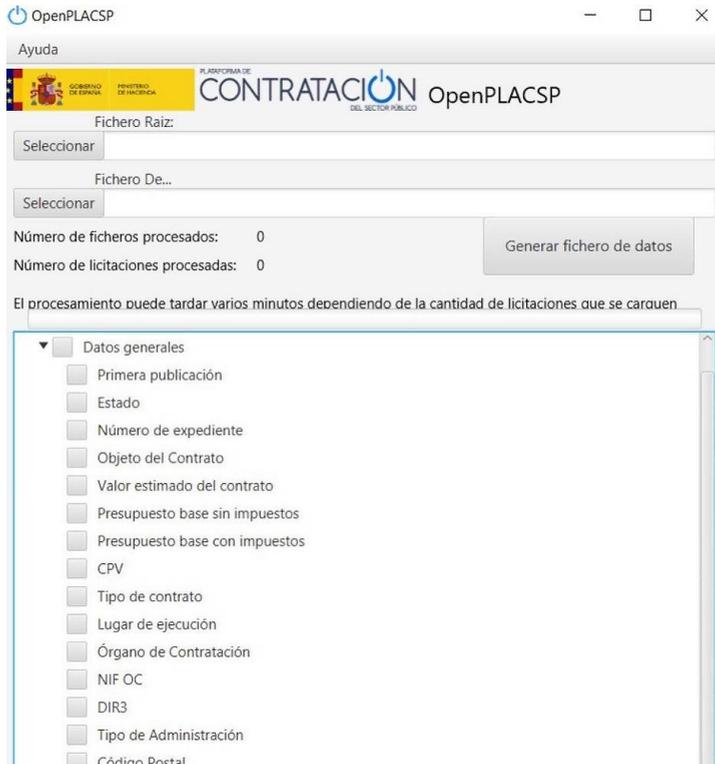


Figura 3.4: Programa OpenPLACSP para extraer licitaciones de Plataforma de Contratación del Sector Público (PLACSP).

En las Tablas 3.2 y 3.3 se recopila una parte de los artículos consultados para llevar a cabo la investigación, sin pretender hacer una revisión exhaustiva de la literatura. Esto se debe a que cada artículo publicado en esta Tesis tiene su correspondiente revisión de la literatura, con más referencias bibliográficas. La mayoría de los artículos consultados se han publicado en los últimos 6 años. No solo en el campo académico, también organismos como la CNMC empiezan a utilizar algoritmos de IA para detectar prácticas anticompetitivas mediante el procesamiento masivo de datos PLACSP¹⁸.

¹⁸. Para más información léase el capítulo “La utilización del big data y la IA para la detección de ilícitos de competencia” [6] y véase el vídeo en el que la Unidad de Inteligencia Económica de la CNMC habla de los algoritmos y herramientas que utilizan: <https://www.youtube.com/watch?v=KQgDnvrRAGA&t=8159s>

Área	Ref.	Título del artículo o libro
Datos en abierto y calidad del dato	[54]	“Data Quality Barriers for Transparency in Public Procurement”
	[39]	<i>The Economic Impact of Open Data Opportunities for value creation in Europe</i>
	[55]	“Modeling of Open Government Data for Public Sector Organizations Using the Potential Theories and Determinants–A Systematic Review”
	[56]	“Open government data portals in the European Union: Considerations, development, and expectations”
	[57]	<i>Open Data Exposed</i>
	[22]	“DIGIWHIST Recommendations for the Implementation of Open Public Procurement Data An Implementer’s Guide”
	[58]	“Open data: Quality over quantity”
	[59]	<i>Datos Abiertos: Guía estratégica para su puesta en marcha Conjuntos de datos mínimos a publicar</i>
	[60]	“Exploring the economic value of open government data”
	[61]	“A systematic review of open government data initiatives”
	[62]	“The exploitation of Business Register data from a public sector information and data protection perspective: A case study”
Innovación y gestión en la contratación	[63]	“AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings”
	[64]	“Innovation and public procurement: Terminology, concepts, and applications”
	[65]	“Project procurement management: A structured literature review”
	[66]	<i>Global Public Procurement Theories and Practices</i>
	[67]	“Research perspectives on public procurement: Content analysis of 14 years of publications in the journal of public procurement”

Tabla 3.2: Recopilación de la literatura relacionada con los datos en abierto y calidad del dato, la innovación y gestión en la contratación.

Las Tablas se han dividido en 4 áreas temáticas:

Área	Ref.	Título del artículo o libro
Forecasting en la contratación	[68]	“Predicting The Number of Bidders in Public Procurement”
	[69]	“Big Data with deep learning for benchmarking profitability performance in project tendering”
	[70]	“ Predicting bid prices by using machine learning methods”
	[71]	“Predicting distresses using deep learning of text segments in annual reports”
	[72]	“Optimized artificial intelligence models for predicting project award price”

Área	Ref.	Título del artículo o libro
Colusión y corrupción	[73]	“Deep learning for detecting bid rigging: Flagging cartel participants based on convolutional neural networks”
	[74]	“Anomaly Detection in Public Procurements using the Open Contracting Data Standard”
	[75]	“Artificial Intelligence Techniques to Detect and Prevent Corruption in Procurement: A Systematic Literature Review”
	[76]	“Transnational machine learning with screens for flagging bid-rigging cartels”
	[77]	“A Machine Learning Approach for Flagging Incomplete Bid-rigging Cartels”
	[78]	“Inferring about fraudulent collusion risk on Brazilian public works contracts in official texts using a Bi-LSTM approach”
	[79]	“Corruption Red Flags in Public Procurement: New Evidence from Italian Calls for Tenders”
	[80]	“Preventing rather than punishing: An early warning model of malfeasance in public procurement”
	[81]	“Machine learning with screens for detecting bid-rigging cartels”
	[82]	“Prediction of Public Procurement Corruption Indices using Machine Learning Methods”
	[84]	“Detecting Fake Suppliers using Deep Image Features”
	[84]	“Predicting Public Procurement Irregularity: An Application of Neural Networks”
	[85]	“Detecting the collusive bidding behavior in below average bid auction”
[86]	“Detecting fraud, corruption, and collusion in international development contracts: The design of a proof-of-concept automated system”	

Tabla 3.3: Recopilación de la literatura relacionada con el forecasting en la contratación y la colusión y corrupción.

- Datos en abierto y calidad del dato. Los datos en abierto gubernamentales y, especialmente, los ligados a la contratación y su calidad, son el pilar para la investigación de esta Tesis.
- Innovación y gestión en la contratación. Artículos que desarrollen estas áreas y que engarzan con la finalidad de esta Tesis.
- Forecasting en la contratación. Algoritmos de ML aplicados a la predicción (estimación) para resolver problemas en el ámbito de los contratos públicos: estimar el número de ofertantes, el importe de adjudicación, etc.
- Colusión y corrupción en la contratación. La colusión (en inglés, collusion o bid-rigging) es un pacto ilegal (coordinado) entre los ofertantes para no competir en la licitación e incrementar su margen de beneficios [87]. La corrupción es el abuso del poder público para obtener beneficios privados [79]. Se recopilan artículos donde la IA y, particularmente, el ML sea el mecanismo para detectar la colusión o corrupción.

3.7. MÉTRICAS DE EVALUACIÓN PARA LOS ALGORITMOS DE ML

Las métricas de evaluación, también llamadas métricas de error, tienen como finalidad evaluar el performance del algoritmo de predicción. Es decir, analizar cómo se comporta la predicción del algoritmo (salida) para el conjunto de datos de entrada. Los problemas que se analizan en esta Tesis son del tipo aprendizaje supervisado: un modelo aprende a relacionar unas variables de entrada y predice, estima, una variable de salida. En particular, hay dos tipos de predicciones o problemas:

- **Problemas de regresión:** se predice un valor numérico. Los artículos de la Tesis que tratan este tipo de problema son:
 - *Public Procurement Announcements in Spain: Regulations, Data Analysis, and Award Price Estimator Using Machine Learning* [10]. Se va a estimar el importe de adjudicación de una licitación gracias al algoritmo de ML llamado Random Forest.
 - *Award Price Estimator for Public Procurement Auctions Using Machine Learning Algorithms: Case Study with Tenders from Spain* [11]. Se va a estimar el importe de adjudicación de una licitación, mejorando el artículo anterior por usarse varios algoritmos, no sólo Random Forest, y obtener el algoritmo que mejor predice (menos error tiene).
- **Problemas de clasificación:** se predice una clase, una categoría, de entre varias previamente definidas. Los artículos de la Tesis que tratan este tipo de problema son:

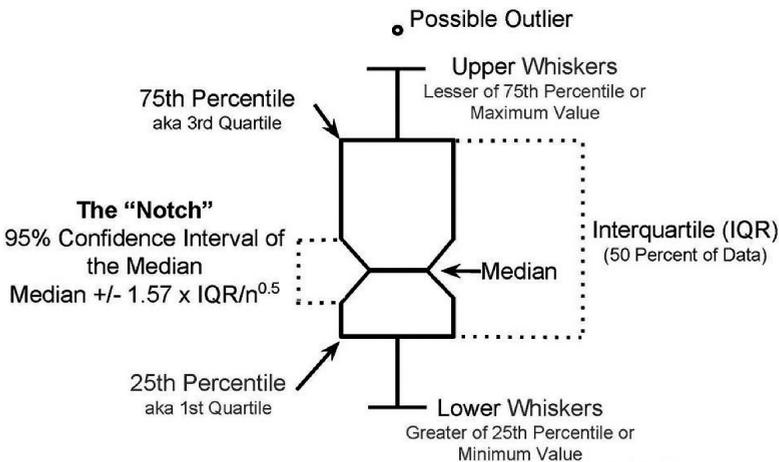


Figura 3.5: Elementos que componen un diagrama de cajas (boxplot).

- *Bidders Recommender for Public Procurement Auctions Using Machine Learning: Data Analysis, Algorithm, and Case Study with Tenders from Spain* [12]. Se va a clasificar el perfil de empresa ganadora de una licitación y, a partir de ella, recomendar un grupo de empresas similares a la estimada.
- *Collusion detection in public procurement auctions with machine learning algorithms* [13]. La colusión se puede simplificar a una clasificación binaria: clase 0 (no hay colusión; es una licitación honesta, competitiva) y clase 1 (sí hay colusión; es una licitación deshonestas, pactaron precios los ofertantes).

En varios de los artículos de la Tesis se va a utilizar el diagrama de caja (denominado boxplot en inglés). Es un método estandarizado para representar gráficamente, de manera compacta, una serie de datos numéricos a través de sus cuartiles, media, mediana y valores atípicos. Ver la Figura 3.5 que identifica sus elementos característicos.

Cada tipo de problema tiene asociado unas métricas de error, las cuales se explican en los dos siguientes subapartados.

3.7.1. Métricas de evaluación para problemas de regresión

Para conocer la precisión de los algoritmos de ML que hacen predicciones de variables continuas, se necesitan formular una serie de métricas de error. De esta manera se podrá saber qué algoritmos estiman mejor, es decir, resuelven mejor el problema de predicción. Sea r_i las observaciones actuales (valores reales), r es la media de las observaciones actuales, p_i las observaciones estimadas y N el número de observaciones.

En las siguientes 10 ecuaciones se define el error absoluto (absolute error, AE), error porcentual absoluto (absolute percentage error, APE), error absoluto medio (mean absolute error, MAE), error absoluto relativo (relative absolute error, RAE), error porcentual absoluto medio (mean absolute percentage error, MAPE), error absoluto mediano (median absolute error, MdAE), error porcentual absoluto mediano (median absolute percentage error, MdAPE), error cuadrático medio (root mean squared error, RMSE), el error cuadrático medio normalizado (normalized root mean squared error, NRMSE) y el coeficiente de determinación¹⁹ (coefficient of determination, R^2).

¹⁹ Es la proporción de la variación de la variable dependiente que es predicha por las variables independientes.

$$AE_i = |r_i - p_i| \quad (3.1)$$

$$APE_i = 100 \left| \frac{AE_i}{r_i} \right| = 100 \left| \frac{r_i - p_i}{r_i} \right| \quad (3.2)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N AE_i = \frac{1}{N} \sum_{i=1}^N |r_i - p_i| \quad (3.3)$$

$$RAE = \frac{\sum_{i=1}^N |r_i - p_i|}{\sum_{i=1}^N |r_i - \bar{r}|} \quad (3.4)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N APE_i = \frac{100}{N} \sum_{i=1}^N \left| \frac{r_i - p_i}{r_i} \right| \quad (3.5)$$

$$MdAE = \frac{1}{N} \text{median}(|r_1 - p_1|, |r_2 - p_2|, \dots, |r_N - p_N|) \quad (3.6)$$

$$MdAPE = \frac{100}{N} \text{median} \left(\left| \frac{r_1 - p_1}{r_1} \right|, \left| \frac{r_2 - p_2}{r_2} \right|, \dots, \left| \frac{r_N - p_N}{r_N} \right| \right) \quad (3.7)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i - p_i)^2} \quad (3.8)$$

$$NRMSE = \frac{RMSE}{\max(r_i) - \min(r_i)} \quad (3.9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N |r_i - p_i|^2}{\sum_{i=1}^N |r_i - \bar{r}|^2} \quad (3.10)$$

3.7.2. Métricas de evaluación para problemas de clasificación

En los problemas de clasificación hay 5 errores importantes para analizar el buen desempeño del algoritmo clasificador: accuracy, precision, recall, balanced accuracy y F1 score. Sea r_i las observaciones actuales (valores reales), p_i las observaciones estimadas, N el número de observaciones y L el número de clases a clasificar ($1 \leq l \leq L$). Los problemas tratados en esta Tesis serán clasificaciones binarias, luego $L = 2$.

El *accuracy* es definido como la proporción de predicciones correctas respecto al total de predicciones hechas:

$$\text{Accuracy} = \frac{1}{N} \sum_{i=1}^N 1(p_i = r_i) \quad (3.11)$$

donde $1(p_i)$ es la función indicador: devuelve 1 si la clase coincide entre p_i y r_i y 0 en otro caso.

El *precision*, también llamado observaciones predictivas positivas, es la habilidad del clasificador para no clasificar una clase como positiva cuando realmente es negativa. El *recall*, también llamada sensibilidad, es la habilidad del clasificador para descubrir todos los valores positivos. Sea r_l el subconjunto de valores positivos de la clase l y p_l el subconjunto de predicciones positivas en la misma clase l :

$$Precision_l = \frac{|r_l \cap p_l|}{|r_l|} \quad (3.12)$$

$$Recall_l = \frac{|r_l \cap p_l|}{|p_l|} \quad (3.13)$$

El *balanced accuracy* se utiliza para evitar desviaciones del *accuracy* en aquellos datos con clases muy desbalanceadas. Es decir, que una de las clases aparece mucho más que otras clases. Su definición es:

$$Balanced\ accuracy = \frac{1}{L} \sum_{l=1}^L recall_l = \frac{1}{L} \sum_{l=1}^L \frac{|r_l \cap p_l|}{|p_l|} \quad (3.14)$$

El *F1 score* es interpretado como la media ponderada del *precision* y *recall*, siendo 1 su mejor valor y 0 el peor:

$$F1\ score = 2 \frac{precision \cdot recall}{precision + recall} \quad (3.15)$$

En particular para nuestra investigación, 0 es la clase de licitaciones competitivas (honestas, donde no hubo alteraciones del importe por los ofertantes) y la clase 1 son licitaciones colusivas (deshonestas, donde los ofertantes formaron un cártel para alterar el importe). Es decir, es una clasificación binaria: clase 1 o 0. Se llama True Positive (TP) a las observaciones de clase 1 correctamente clasificadas como 1 y True Negative (TN) a las de clase 0 también correctamente clasificadas. El False Positive (FP) son las observaciones de clase 0 que se clasifican incorrectamente, asignándoles la clase 1. Análogamente, el False Negative (FN) son las observaciones de clase 1 que se clasifican incorrectamente, asignándoles la clase 0. Lógicamente, el número total de observaciones debe ser la suma de TP + TN + FP + FN. Estos valores se pueden representar en lo que se llama matriz de confusión (ver Figura 3.6).

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Figura 3.6: Matriz de confusión para la clasificación binaria.

Las fórmulas anteriores se simplifican para la clasificación binaria:

$$Accuracy = \frac{TP + TN}{N} \quad (3.16)$$

$$Precision = \frac{TP}{TP + FP} \quad (3.17)$$

$$Recall = \frac{TP}{TP + FN} \quad (3.18)$$

$$Balanced\ accuracy = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \quad (3.19)$$

CAPÍTULO CUARTO. ARTÍCULOS DE LA INVESTIGACIÓN

4.1. INTRODUCCIÓN

Hay pocos investigadores y artículos académicos que se dediquen a aplicar la ciencia de datos, el análisis masivo de datos y la IA a la contratación pública, ni en España ni en otros países. Esto se debe fundamentalmente a que se necesitan grandes volúmenes de datos, no disponiéndose de ellos hasta hace pocos años, gracias a la publicación masiva de contratos en las plataformas de contratación. Contratos públicos de cualquier sector, importe y ámbito geográfico. Tradicionalmente, los estudios similares a los de esta Tesis se centraban solamente en el sector de la construcción pública e ingeniería civil, debido a que son los de mayor importancia en la contratación por su alto coste económico y gran impacto para la sociedad.

Este capítulo está organizado en los siguientes apartados, cada uno correspondiéndose con una investigación en particular. Se va a explicar de la manera más sencilla posible los artículos del compendio de publicaciones de la Tesis. Es decir, que no sea demasiado complejo para que un lector no técnico, no experto en IA, pueda entenderlo. Así se consigue abrir la Tesis a investigadores de otros ámbitos, como el jurídico o económico. El lector técnico, especializado en IA y programación, debería leer los artículos (anexos al final de la Tesis) para profundizar en cómo se resuelve el problema, mediante qué técnicas cuantitativas y de ML.

- **Licitaciones en España: regulación, análisis de datos y estimador del importe de adjudicación usando ML [10].** Primer artículo del compendio de la Tesis cuyo objetivo es introducir la analítica de datos a la contratación y crear un estimador del precio de adjudicación con un algoritmo de ML.

- **Estimador del importe de adjudicación de licitaciones usando ML: caso de estudio con licitaciones de España** [11]. Segundo artículo y extensión del anterior, consiguiendo un mejor estimador del importe de adjudicación. Se testaron varios algoritmos de IA, batiendo alguno de ellos al algoritmo anterior.
- **Recomendador de licitadores usando ML: análisis de datos, algoritmo y caso de estudio con licitaciones de España** [12]. El tercer artículo de la Tesis crea un buscador de empresas que puedan llevar a cabo una licitación dada. Gracias al histórico de licitaciones y a la fuente de datos empresariales, se ha entrenado un algoritmo de ML que permite a un usuario introducir una licitación y obtener qué empresas son las más capacitadas, a priori, para acometer el trabajo.
- **Detección de colusión en licitaciones aplicando algoritmos de ML** [13]. Se presenta el grave problema de la colusión en la contratación pública, es decir, aquellas empresas que forman un cártel para pactar ofertas económicas en las licitaciones. Se comparan 11 algoritmos de ML para detectar si una licitación es competitiva (los ofertantes no pactaron precios) o colusiva.

Además, se explicará una **aplicación informática para detectar licitaciones irregulares**. Tiene como objetivo detectar licitaciones donde ha podido existir actuaciones ilegales, fraudulentas. Se detectaron varias licitaciones de este tipo en España y Andalucía.

Las *fuentes de datos* usadas en los artículos son de dos tipos:

- **Datos de contratación**. Se ha utilizado PLACSP en todos los artículos salvo en el último artículo (detección de colusión). Además, en uno de los artículos se han utilizado licitaciones europeas obtenidas de TED. Estas dos fuentes de información son datos en abierto (públicas y gratuitas). Por otro lado, en el último artículo sobre colusión se han utilizado 6 datasets de 5 países distintos, de acceso restringido en su mayoría (no son datos en abierto), solicitándolos directamente a organizaciones públicas (Autoridades de la Competencia y Policía) o a investigadores.
- **Datos empresariales**. Se ha conseguido la información más relevante de las cuentas anuales de 1,3 millones de empresas españolas, principalmente entre los años de 2014 a 2019, provenientes del Registro Mercantil. Estos datos son de acceso público pero de pago.

Los artículos se han programado en el lenguaje de programación Python [3, 40, 41] con la librería de ML llamada *Scikit-learn* [88, 89]. Para el segundo artículo, que emplea redes neuronales, se ha utilizado la plataforma de ML llamada WEKA [90, 91] y Tensorflow [92].

Todos los artículos son de publicación open access. Así se facilita su lectura y difusión entre los distintos involucrados (directa o indirectamente) en la contratación, como son las AA.PP., organizaciones públicas con competencias en contratación, operadores económicos, investigadores, ciudadanos, etc. Muchos de estos involucrados son ajenos a la Academia y necesitan un acceso libre y gratuito. Es más, gran parte de los investigadores interesados en la contratación (juristas, economistas, politólogos, etc) no tienen suscripciones a las revistas del ámbito científico-tecnológico donde se han publicado estos artículos. La investigación que se va a exponer son ejemplos prácticos, reales, que aportan valor a los organismos públicos que gestionan, controlan y velan por la contratación. Además, los operadores económicos que son contratados por las AA.PP. también obtendrían ventajas y beneficios al usar estas herramientas innovadoras.

4.2. LICITACIONES EN ESPAÑA: REGULACIÓN, ANÁLISIS DE DATOS Y ESTIMADOR DEL IMPORTE DE ADJUDICACIÓN USANDO ML

Este artículo tiene por título original en inglés “*Public Procurement Announcements in Spain: Regulations, Data Analysis, and Award Price Estimator Using Machine Learning*” [10]. Es el primer artículo del compendio de publicaciones de la Tesis y tiene como objetivo principal hacer un estimador (predictor) del importe de adjudicación de una licitación. Es decir, predecir cuál es el precio de mercado, la oferta económica de la empresa ganadora del contrato. Esto tiene gran importancia porque las AA.PP., al publicar una licitación, estiman el importe de licitación a veces sin mucha información de mercado, errando en su estimación inicial. Por eso hay diferencias (más o menos significativas) entre el importe de licitación y el de adjudicación. A esa diferencia se le suele llamar comúnmente “*baja*” en el argot de la contratación.

Como es un artículo de una temática novel (ciencia de datos aplicado a las licitaciones), se hace una presentación de la legislación de la contratación pública (más detallado en esta Tesis en el apartado correspondiente del capítulo 3), se describe PLACSP, los campos de las licitaciones que se han utilizado y cómo se han procesado los datos.

La metodología del artículo se muestra en la Figura 4.1, así como el software utilizado (Python, sus librerías típicas de ciencia de datos y MySQL). Como fuente de datos se utiliza la española PLACSP y la europea TED. Ambas fuentes se utilizan de manera independiente, por separado, para dotar al artículo de mayor generalidad, verificándose que se puede crear un estimador del importe de adjudicación tanto a nivel español como europeo. Después se describe cómo se han procesado el conjunto de datos (extracción, reducción, limpieza, transformación y filtrado) y el análisis de datos realizado (análisis estadístico, matriz de correlación y dispersión, histogramas, etc.). Finalmente, se crea

el estimador del importe de adjudicación mediante un algoritmo de ML, llamado Random Forest, que se evalúa y ajusta gracias a las métricas de evaluación del error (MAPE, MdAPE, RMSE, etc.). En este caso, el error es la diferencia entre la estimación (valor predicho) del importe de adjudicación y el valor real adjudicado.

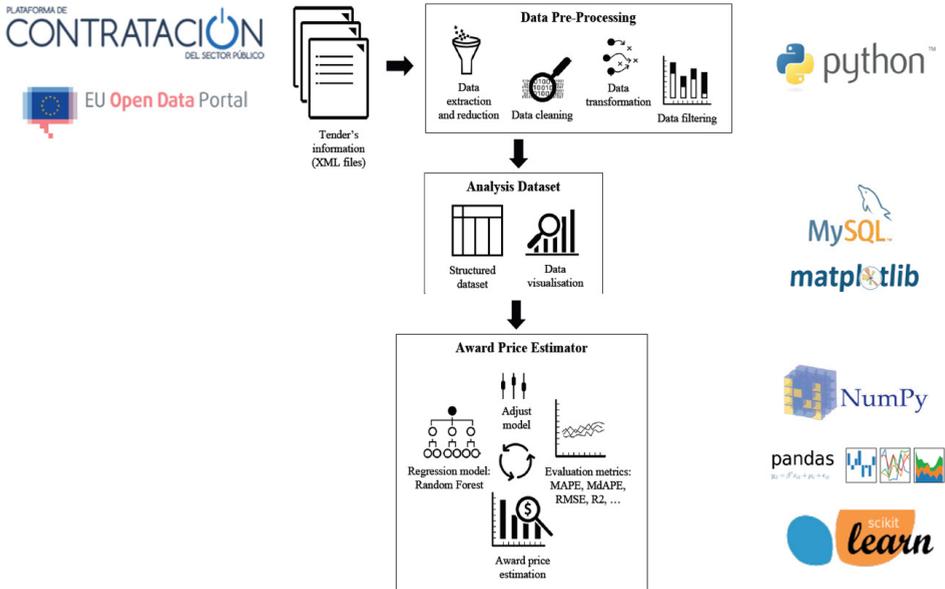


Figura 4.1: Flujograma del análisis de datos de contratación y el estimador del precio de adjudicación. A la derecha, las librerías software utilizadas.

Se puede crear el estimador del importe de adjudicación gracias a que el algoritmo de ML se entrena con un gran volumen de licitaciones, incluyendo sus importes de licitación y adjudicación, y es capaz de estimar el precio de mercado para una licitación dada. Para el caso de España, se entrenó con casi 60.000 licitaciones multisectoriales publicadas entre 2012 y 2018. Para el caso de Europa, se entrenó con aproximadamente 41.500 licitaciones multisectoriales publicadas en 2017 de 8 países europeos.

En la Tabla 4.1 se resumen las principales magnitudes de los datasets de España y Europa, las variables de entrada para el algoritmo Random Forest y los resultados que obtiene el estimador del importe de adjudicación. Para el caso de España, el MdAPE (error porcentual absoluto mediano) entre el importe adjudicación y de licitación es del 11,84% y, sin embargo, el MdAPE entre el importe de adjudicación y la predicción es del 9,26%. Es decir, el estimador es más fiable por tener un -2,58% menos de error. Si lo analizamos desde la perspectiva del MAPE (error porcentual absoluto medio), el estimador sigue siendo más fiable por tener un -11,19% menos de error. Para el caso de Europa, el estimador obtiene mejor resultado en media (el MAPE es menor)

y en mediana es un poco peor (el MdAPE es ligeramente superior). La causa de que el estimador obtenga peores resultados con licitaciones europeas es porque la “baja” es muy pequeña. Es decir, en Europa no hay grandes diferencias entre el importe de licitación y adjudicación, hecho que sí sucede en España. Esto se observa en el MdAPE, el de Europa es casi 3 veces superior al de España. Este suceso es sorprendente y debería de analizarse con más licitaciones, de otros años y más países.

	España	Europa
Nº licitaciones	58.337	41.556
Ubicación	Toda España	Francia, Alemania, Italia, Hungría, Letonia, Croacia, Eslovenia y Bulgaria
Fechas licitaciones	2012-2018	2017
Nº órganos de contratación	3.544	6.163
Nº empresas adjudicatarias	17.305	19.100
Algoritmo de predicción	Random Forest	Random Forest
VARIABLES DE ENTRADA	Importe licitación, Fecha, Duración, Organismo licitador, CPV, CPV Agregado, Código procedimiento, Código tipo, Código subtipo, Código urgencia, Código Postal, CC.AA., Provincia y Municipio	Importe licitación, Fecha, Organismo licitador, Actividad principal, CPV, CPV Agregado, Código procedimiento, Código tipo, Código Postal y País
Variable salida (predicción)	Importe adjudicación	Importe adjudicación
MdAPE entre importe de adjudicación y licitación (azul) y MdAPE entre adjudicación y predicción (gris)	11,84% 9,26% (-2,58% menos de error)	4,17% 6,48% (+2,31% más de error)
MAPE entre importe de adjudicación y licitación (azul) y MAPE entre adjudicación y predicción (gris)	39,79% 28,60% (-11,19% menos de error)	27,49% 23,57% (-3,92% menos de error)

Tabla 4.1: Datasets utilizados y métricas de error del estimador del precio de adjudicación.

Para comprender de manera gráfica cómo predice el algoritmo, observar la Figura 4.2 (datos de España). En la gráfica superior, se muestra el importe de licitación (eje x) y el importe de adjudicación (eje y). Si las AA.PP. españolas acertasen totalmente en los importes de adjudicación que dan cuando se publica la licitación, esta gráfica tendría que ser una

línea recta a 45 grados del eje x. Es decir, las AA.PP. acertarían siempre y el importe de licitación coincidiría con el precio de mercado (el importe finalmente adjudicado). Sin embargo, se aprecia como hay una nube de puntos, habiendo significativas diferencias entre ambos importes. En la gráfica inferior, se muestra la estimación del importe de adjudicación (eje x) y el importe real de adjudicación (eje y). Se observa cómo sigue existiendo una nube de puntos porque, como ya vimos, el estimador no es perfecto, tiene error. Sin embargo, los puntos están más concentrados respecto a la línea recta descrita anteriormente que sería la ideal, donde no habría error.

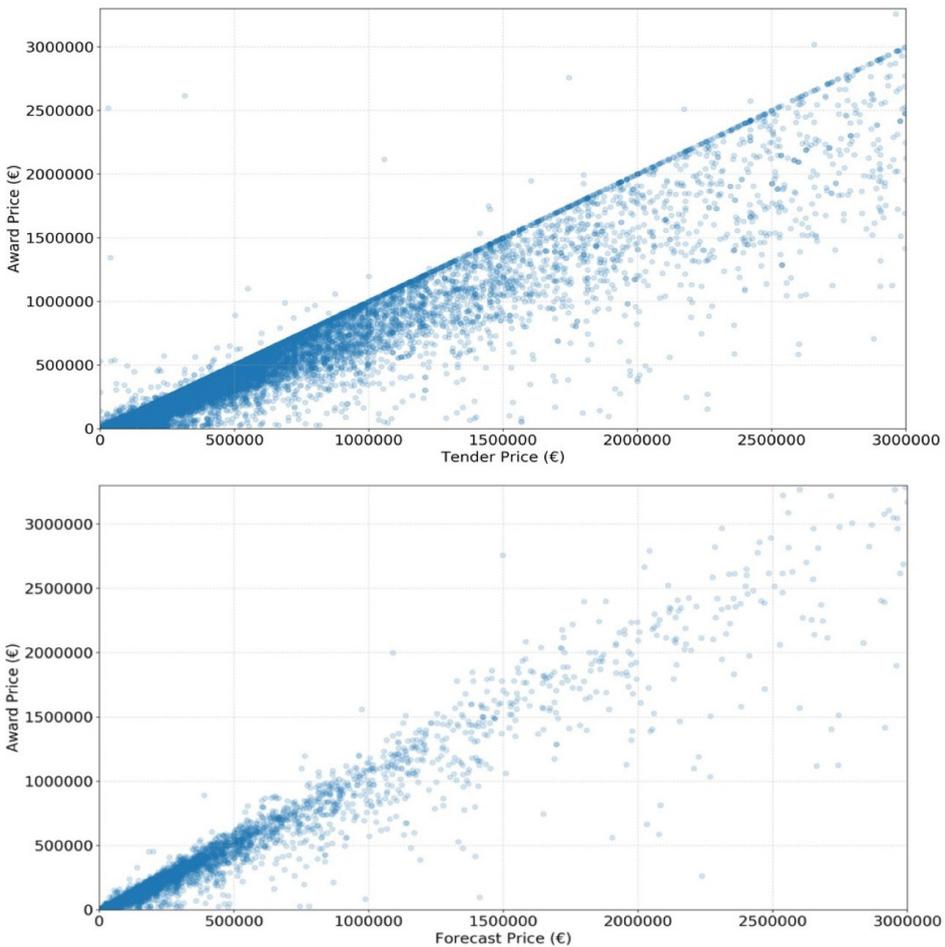


Figura 4.2: En la gráfica superior se compara el importe de licitación (eje x) frente al importe de adjudicación (eje y). En la inferior, la predicción del importe de adjudicación (eje x) frente al importe de adjudicación (eje y). Datos de España.

El algoritmo Random Forest se usará también en el resto de artículos de la Tesis. Así que, por su importancia, se hace aquí un breve resumen (en los artículos está su pseudocódigo y más información). Random forest, introducido por Breiman [93] en 2001, es un método de aprendizaje de conjuntos para regresión o clasificación que opera mediante la construcción de una multitud de árboles de decisión en el entrenamiento y la salida es la moda de las clases (si es un problema de clasificación) o la predicción media (si es un problema de regresión) de los árboles individuales. Los dos primeros artículos de la Tesis son problemas de regresión (estimar un valor numérico) y los dos últimos artículos son problemas de clasificación. Es un popular algoritmo de ML que ofrece excelente performance [94, 95], no overfitting [96], una versatilidad para aplicarse a problemas reales con gran volumen de datos y con diferentes tipos de datos [94, 97]. En este artículo (y en los demás de la Tesis) se coge un 80% de los datos como dataset de entrenamiento y un 20% para la validación (test) del algoritmo.

4.3. ESTIMADOR DEL IMPORTE DE ADJUDICACIÓN DE LICITACIONES USANDO ML: CASO DE ESTUDIO CON LICITACIONES DE ESPAÑA

Este artículo tiene por título original en inglés “*Award Price Estimator for Public Procurement Auctions Using Machine Learning Algorithms: Case Study with Tenders from Spain*” [11]. Es una extensión del anterior artículo, utilizando el mismo conjunto de datos de España (el de Europa no se utiliza), cuyo objetivo es utilizar más algoritmos de ML para evaluar cuál resuelve mejor el problema de predicción del importe de adjudicación. Por tanto, la metodología es análoga al artículo anterior, no teniendo que hacer las primeras fases (procesado el conjunto de datos y análisis) por ser el mismo dataset de España.

Para la experimentación computacional de los algoritmos se ha utilizado WEKA [90, 91], la plataforma de ML desarrollada por U. de Waikako que da soporte a un gran número de algoritmos de ML. Se han utilizado 4 tipos de algoritmos de ML:

- Random Forest. Se toma como modelo base para establecer la comparación por ser el utilizado en el primer artículo.
- Regresión lineal. Relaciona de manera lineal (una recta) las variables de entrada y la variable de salida (predicción)²⁰. Es un modelo clásico, siendo la primera referencia documentada por el matemático Legendre en 1805.
- Regresión isotónica (o monótona). Hace una aproximación de una función no decreciente (monótona) minimizando el error cuadrático medio (MSE) en los datos de entrenamiento²¹.

²⁰ Consultar el artículo para su definición matemática.

²¹ Consultar el artículo para su definición matemática.

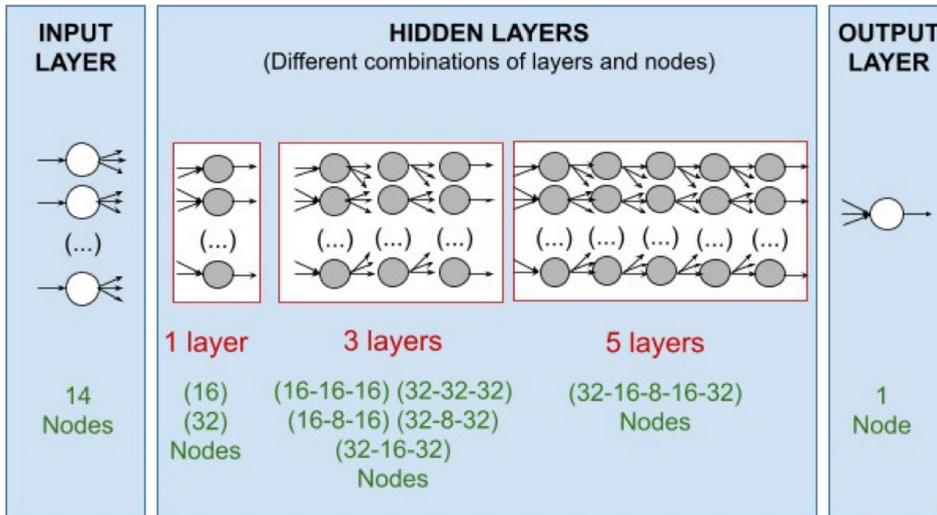


Figura 4.3: Arquitectura de las capas y nodos de las redes neuronales artificiales (ANN) utilizadas.

- Redes neuronales artificiales. En inglés se denominan Artificial Neural Networks (ANNs) y forma parte del subcampo del ML llamado deep learning (aprendizaje profundo). Una ANN es un modelo computacional inspirado en las redes neuronales biológicas. Consiste en una colección de nodos (neuronas artificiales) organizados en capas interconectadas (capa de entrada, capas ocultas y capa de salida). Los parámetros del modelo son los pesos y sesgos asociados a las conexiones. La información se procesa desde la capa de entrada a la capa de salida. El proceso de aprendizaje se basa en minimizar una función de coste (conocida como función de pérdida) que evalúa el error de la ANN para resolver el problema.

Se han probado muchas configuraciones distintas de ANNs. Por un lado, se ha utilizado el MLP (Multi-layer perceptron) [52]. Por otro lado, se han definido 8 arquitecturas distintas de capas y nodos (ver Figura 4.3). Cada configuración viene definida por: número de capas y nodos, función de activación (rectified linear unit (ReLU) o scaled exponential linear unit (SeLU)), función de pérdida (MAE o MSE) y optimizador (Adam, Adamax o Adagrad). La combinación de las 8 arquitecturas diferentes, 2 funciones de activación, 2 funciones de pérdida y 3 optimizadores ha dado lugar a 96 configuraciones distintas de ANN. Las arquitecturas de ANNs se han optimizado usando Tensorflow [92]. Tras la validación de las ANNs, se han seleccionado 4 configuraciones diferentes (ANN1, ANN2, ANN3 y ANN4) que tenían un error más pequeño.

Los resultados de la experimentación realizada (descrita en detalle en el artículo) se muestran en la Tabla 4.2, resaltándose en negrita qué algoritmo tiene el mejor error (el más pequeño) para cada uno de las 4 métricas de error evaluadas (MAE, RMSE, RAE y RRSE). En realidad, son dos errores distintos (MAE y RMSE) y sus respectivos errores relativos (RAE y RRSE). Por eso quien obtenga el mejor MAE, obtendrá consecuentemente el mejor RAE. Análogamente ocurre para el RMSE y RRSE. Las redes neuronales (ANNs) son las que obtiene los mejores errores, en particular el MLP (mejor RMSE y RRSE) y el ANN2 (mejor MAE y RAE).

	MAE	RMSE	RAE	RRSE
Random Forest (baseline model)	179,247.80€	6,621,784.24€	31.11%	74.86%
Linear regression	228,491.36€	15,535,231.61€	39.66%	175.63%
Isotonic regression	136,971.39€	5,648,693.54€	23.76%	63.86%
ANN (MLP)	270,953.50€	1,974,981.24€	47.03%	22.33%
ANN1	140,763.27€	7,416,004.50€	23.03%	83.84%
ANN2	123,570.91€	5,110,687.50€	20.22%	57.78%
ANN3	157,181.00€	9,543,883.00€	25.71%	107.90%
ANN4	124,035.82€	3,304,259.20€	20.29%	37.36%

Tabla 4.2: Métricas de error del estimador del precio de adjudicación para los algoritmos de ML.

4.4. RECOMENDADOR DE LICITADORES USANDO ML: ANÁLISIS DE DATOS, ALGORITMO Y CASO DE ESTUDIO CON LICITACIONES DE ESPAÑA

Este artículo tiene por título original en inglés “*Bidders Recommender for Public Procurement Auctions Using Machine Learning: Data Analysis, Algorithm, and Case Study with Tenders from Spain*” [12]. Se ha creado un buscador de empresas que puedan llevar a cabo un contrato. Es decir, el buscador (un algoritmo desarrollado ad hoc) recomienda un grupo de empresas en base a la información del contrato: órgano de contratación, importe de licitación, código CPV (Common Public Vocabulary), tipo de procedimiento y tramitación, fecha de publicación, etc.

Se han utilizado dos fuentes de datos: licitaciones y empresas. Por un lado, se utilizaron 102.000 licitaciones (de cualquier naturaleza) de PLACSP publicadas entre 2014 y 2020. Por otro lado, se utilizaron 1,3 millones de empresas (ubicación, facturación anual, EBITDA número de empleados, clasificación nacional de actividades económicas, etc.) obtenidas de las cuentas anuales presentadas en el Registro Mercantil de España. En el artículo se describen ambos datasets en sendas tablas de manera detallada. La metodología para extraer y manejar los datos de PLACSP se ha descrito en el primer artículo. Para los datos empresariales, no se ha tenido que hacer ninguna metodología por haberse obtenido de una base de datos con toda la información ya estructurada y validada.



Figura 4.4: Funcionamiento del recomendador de licitadores (buscador de empresas) mediante dos capturas de pantalla de la aplicación web.

Los datos de la PLACSP son de acceso libre y gratuito, como ya se ha mencionado. Sin embargo, actualmente los datos empresariales del Registro Mercantil son de acceso libre pero de pago. Esto supone un handicap para poder desarrollar investigaciones o herramientas en este campo. Algunos papers han propuesto herramientas similares a ésta pero sólo analizando las licitaciones [98, 99], no incluyendo también las características de los proveedores.

Además, se ha desarrollado una aplicación web (ver Figura 4.4) para demostrar de manera práctica, real, el algoritmo de búsqueda planteado en este artículo. A la izquierda de la Figura se muestra el formulario con los campos de la licitación a rellenar y en la parte derecha se arrojan las empresas recomendadas por el buscador. Esta aplicación web fue la ganadora en 2020 del premio [Open Data](#) del Gobierno Vasco²² (dotado con 10.000€). Además, actualmente se encuentra entre los candidatos finalistas al premio [EU Datathon 2022](#), organizado por la Comisión Europea (CE), siendo la temática de esta edición utilizar los datos en abierto de la contratación pública.

El funcionamiento del buscador se describe en la Figura 4.5. El algoritmo que estima la empresa ganadora de la licitación es un Random Forest de tipo clasificación (más detalle en el artículo). La fuente de datos de licitaciones se utiliza solamente para entrenar el modelo de predicción pero la fuente de datos empresarial se utiliza también en fase de ejecución, para buscar las empresas.

²² Concesión en <https://euskadi.eus/y22-bopv/es/bopv2/datos/2020/12/2005583a.shtml>

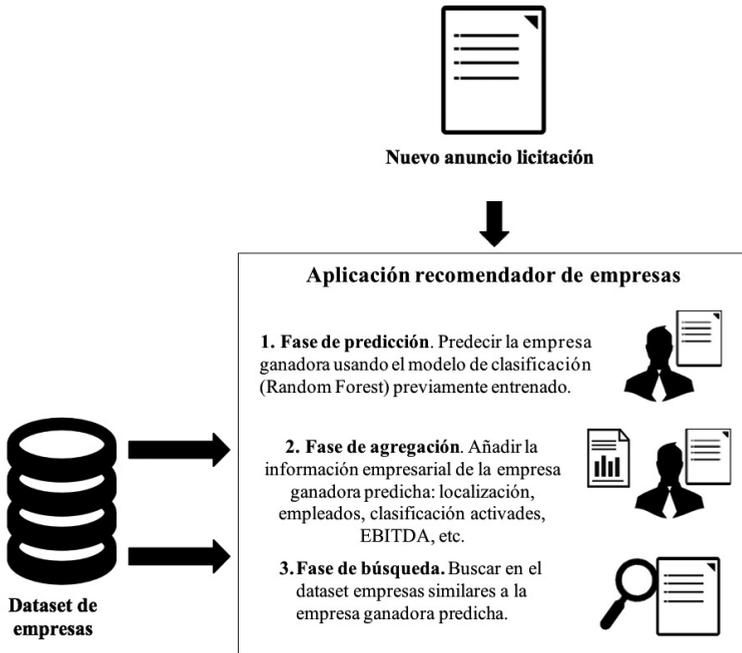


Figura 4.5: Flujograma del funcionamiento del recomendador de licitadores (buscador de empresas) para una nueva licitación de entrada.

En la fase de búsqueda (3), se buscan empresas que sean similares a la empresa predicha (forecast company). Esta similitud se basa en una serie de reglas (ecuaciones) utilizando la facturación, EBIT, EBITDA, número de empleados, distancia entre la ubicación de la licitación y la sede de la empresa (el factor geográfico es determinante en la contratación) y una similitud los códigos de actividad empresarial (se utilizarán 4 tipos distintos de clasificaciones: NACE2, IAE, SIC y NAICS). Dichas magnitudes tienen que ser mayor o igual a los de forecast company ponderados por una constante F_x cuyo valor está comprendido entre 0 y 1.

$$Operating_income_{company} \geq F_{OI} \cdot Operating_income_{forecast\ company} \quad (4.1)$$

$$EBIT_{company} \geq F_{EBIT} \cdot EBIT_{forecast\ company} \quad (4.2)$$

$$EBITDA_{company} \geq F_{EBITDA} \cdot EBITDA_{forecast\ company} \quad (4.3)$$

$$Employees_{company} \geq F_E \cdot Employees_{forecast\ company} \quad (4.4)$$

$$Distance_{tender-company} \geq F_D \cdot Distance_{tender-forecast\ company} \quad (4.5)$$

$$\sum_{i=1}^C 1[Code_{company} = Code_{forecast\ company}] \geq F_{CEA} \cdot C \quad (4.6)$$

donde $1[Code]$ es la función indicador (devuelve 1 si coinciden los códigos y si no devuelve 0), C es el número total de códigos de la forecast company y $Code$ es el código de clasificación de actividades económicas de la compañía (entre los 4 tipos distintos de los que se dispone).

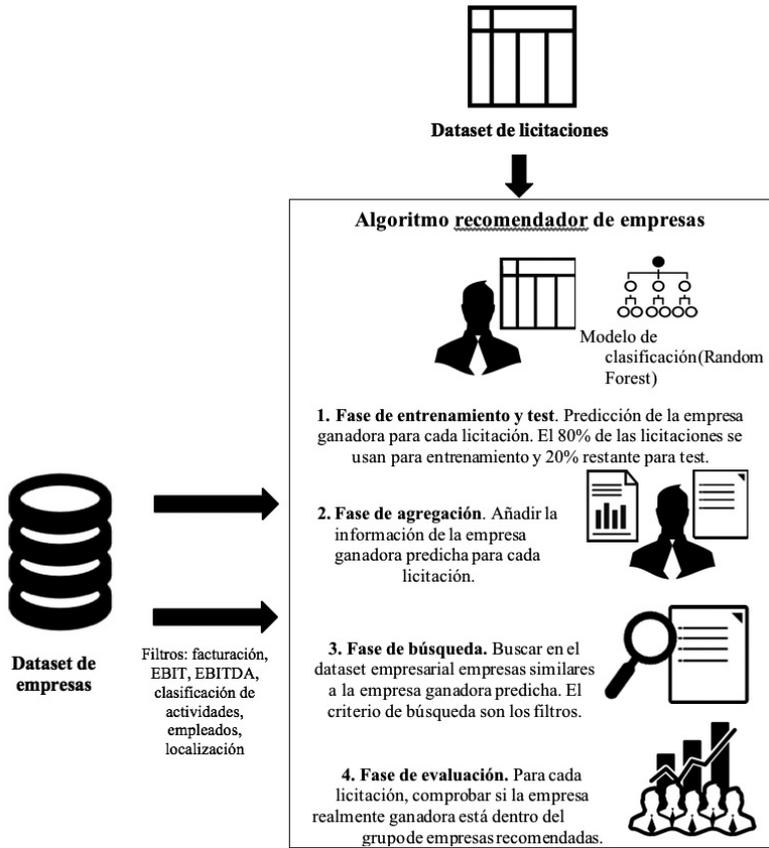


Figura 4.6: Flujograma para crear el algoritmo de ML recomendador de empresas.

Es necesario entrenar el modelo para que el buscador esté operativo. Para ello, se muestra en la Figura 4.6 el flujograma para entrenar dicho algoritmo Random Forest de clasificación. Finalmente, se evalúa el performance del algoritmo con un 20% de las licitaciones.

Los resultados han sido exitosos. La Tabla 4.3 muestra la batería de pruebas realizadas: dos escenarios y 5 configuraciones distintas de las constantes F_x (de menos a más re restrictiva). Por ejemplo, para el escenario 2 y el $Accuracy_{n=5}$, la empresa ganadora está en el grupo de 5 empresas recomendadas en el 31,58% de las veces. Para el escenario 2 dicho valor cae al 23,12%, influenciado seguramente por la dificultad en predecir las licitaciones

más recientes que las más antiguas. Por tanto, es una aplicación que ejemplifica la utilidad que tienen este tipo de herramientas de IA en la contratación. Especialmente para los órganos de contratación que redactan y gestionan contratos de tipo negociado (con o sin publicidad) y necesitan buscar empresas que puedan acometer licitaciones pero también para las propias empresas licitadoras que pueden sondear qué competencia tienen en el mercado.

Description	Different bidders recommender settings					
	Very low	Low	Medium	High	Very high	
Bidders recommender factors for the settings	F_{OI} : operating income factor	0.25	0.5	0.65	0.75	1.0
	F_{EBIT} : EBIT factor	0.25	0.5	0.65	0.75	1.0
	F_{EBITDA} : EBITDA factor	0.25	0.5	0.65	0.75	1.0
	F_E : employees factor	0.15	0.25	0.25	0.35	0.45
	F_{CEA} : classification economic activities factor	0.125	0.15	0.14	0.175	0.2
	F_D : distance tender-company factor	1.6	1.4	1.4	1.2	1
Results of scenario 1: testing subset is the 20% of the dataset randomly chosen	Accuracy _{n=1} : winner company is the forecast company			17.07%		
	Accuracy _{n=5} : winner company is within the top 5 forecast companies			31.58%		
	Accuracy _{n=M} : winner company is within the recommended companies group	38.52%	36.20%	35.92%	34.04%	33.25%
	Mean and median number of the recommended companies of each tender	877.43; 86	469.69; 35	430.48; 31	226.07; 11	145.97; 9
		Accuracy _{n=1} : winner company is the forecast company			10.25%	
Results of scenario 2: testing subset is the last 20% of the dataset ordered by tender's date	Accuracy _{n=5} : winner company is within the top 5 forecast companies			23.12%		
	Accuracy _{n=M} : winner company is within the recommended companies group	30.52%	28.00%	27.73%	25.55%	24.79%
	Mean and median number of the recommended companies of each tender	900.64; 95	470.41; 37	430.33; 33	210.92; 11	132.10; 9

Tabla 4.3: Métricas de error del recomendador de licitadores para distintas configuraciones en dos escenarios diferentes.

4.5. DETECCIÓN DE COLUSIÓN EN LICITACIONES APLICANDO ALGORITMOS DE ML

Este artículo tiene por título original en inglés “*Collusion detection in public procurement auctions with machine learning algorithms*” [13]. La colusión es una práctica ilegal mediante la cual algunas empresas competidoras acuerdan en secreto las ofertas económicas que presentarán a una licitación. La colusión es un fenómeno generalizado en la contratación pública de todo el mundo (pudiendo también aparecer en la contratación privada). Estas prácticas colusorias suelen adoptar la forma de aumentos de precios coordinados (no competitivos) establecidos entre las empresas (denominados cárteles) [100]. Por tanto, se socavan los beneficios del mercado competitivo, malgastando el dinero de los contribuyentes. Generalmente, los órganos de contratación no pueden identificar ofertas no competitivas, es decir, adjudicando contratos a precios más altos de los que habrían tenido sin colusión.

Description	Brazil	Italy	Japan	Swiss-Ticino	Swiss-SG&GR	US
Scope	Oil infrastructure projects	Road construction	Building constr. and civil eng.	Road construction	Road construction and civil engineering	School milk market
Time period	2002-2013	2000-2003	2003-2007	1999-2006	14 years (over 2005)	1980-1990
N° auctions	101	278	1080	224	4344	3754
N° bids	683	20,286	13,515	1629	21,231	7004
Awarding criteria	Lowest bid	Average Bid Method	Lowest bid	Lowest bid	Lowest bid	Lowest bid
Avg. n° of bids per auction	6.76	72.97	12.51	7.27	4.89	1.91

Tabla 4.4: Descripción de los datasets (licitaciones competitivas y colusivas) de Brasil, Italia, Japón, Suiza-Ticino, Suiza-SG&GR y USA.

Se van a comparar 11 algoritmos de ML para detectar colusión utilizando datasets obtenidos de Brasil, Italia, Japón, Suiza (de dos regiones llamadas Ticino y St Gallen & Graubünden) y Estados Unidos (ver la Tabla resumen 4.4 y para más detalle consultar el propio artículo). Son aproximadamente 10.000 licitaciones de todo el mundo, de distinta naturaleza (obras públicas, infraestructuras civiles, suministro de leche a escuelas, etc.), de distintos periodos temporales y provenientes de fuentes oficiales por lo que la fiabilidad y calidad del dato es excelente²³. Por tanto, es un gran conjunto para experimentar en la detección de colusión aplicando los algoritmos de ML, siendo éste el primer artículo que reúne juntos tantos datasets. Nótese que son extremadamente difíciles de conseguir por no ser información accesible al público. Generalmente está restringida a las Autoridades de la Competencia u otras instituciones públicas que investigan prácticas anticompetitivas, demostrando la existencia de los cárteles ante los tribunales de justicia.

La metodología del artículo se muestra en la Figura 4.7. Las dos primeras fases son las tareas típicas de ciencia de datos y la tercera fase (calcular las variables de cribado) es propia para la colusión. Las variables de cribado (screening variables o, simplemente, screens) son variables secundarias o derivadas, calculadas a partir de las ofertas económicas, que ayudan a identificar licitaciones potencialmente colusivas [81]. Estas variables han sido propuestas por diferentes investigadores en el campo de la detección de colusión [87, 76, 77, 101, 102]. Se han utilizado las siguientes screens²⁴ (denominaciones en inglés): Coefficient of Variation (CV_i), Spread (SPD_i), la diferencia relativa entre las dos licitaciones más bajas en la licitación ($DIFFP_i$), la Relative Distance (RD_i), el Skewness ($SKEW_i$), el Excess Kurtosis ($KURT_i$) y el Kolmogorov-Smirnov test ($KSTEST_i$).

²³ Agradecer al Dr. David Imhof (investigador y trabajador en la [Autoridad Suiza de la Competencia](#)) y al Dr. Regis Signor (investigador e Inspector en la [Policía Federal de Brasil](#)) por facilitar varios de los datasets. Además, reconocer al Dr. Imhof por sus artículos pioneros en este campo (detección de colusión mediante ML) y al Dr. Signor por haber colaborado en este artículo.

²⁴ Consultar el artículo para sus definiciones matemáticas.

La detección de colusión se mide bajo diferentes escenarios y cantidad de información de entrada. Se supone que cada licitación se puede clasificar como “colusión” o “competitiva”. Por tanto, el algoritmo de ML tiene que realizar una clasificación binaria para cada licitación. Se van a utilizar y comparar los siguientes 11 algoritmos (denominación en inglés):

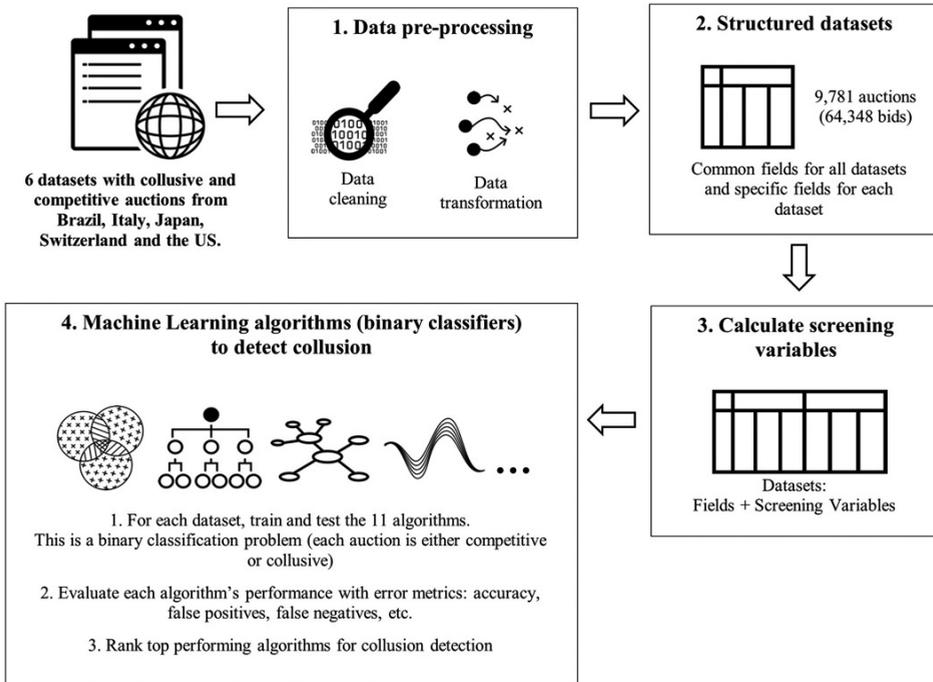


Figura 4.7: Flujograma para la detección de colusión mediante ML.

- Linear models: SGD (Stochastic Gradient Descent) [103].
- Ensemble methods: Extra Trees (Extremely Randomized Trees) [104], Random Forest [93], Ada Boost [105] and Gradient Boosting [106].
- Support Vector Machines: SVC (C-Support Vector Classification) [107].
- Nearest Neighbors: K Neighbors [108].
- Artificial Neural Network: MLP (Multi-Layer Perceptron) [52].
- Naive Bayes: Bernoulli Naive Bayes and Gaussian Naive Bayes [52].
- Gaussian Process [109].

La detección de colusión se va a testear bajo 4 escenarios distintos de información de entrada (settings). Es decir, el algoritmo de ML dispondrá de 4 configuraciones distintas de variables (campos) de entrada para poder hacer la detección:

- Setting 1. Se utilizan todos los campos disponibles.
- Setting 2. Se utilizan todos los campos disponibles más los screens.
- Setting 3. Se utilizan los campos comunes a todos los datasets.
- Setting 4. Se utilizan los campos comunes a todos los datasets más los screens.

Topic	Description	Datasets							All datasets
		Brazil	Italy	Japan	Swiss - Ticino	Swiss - SG&GR	US		
Fields	Common fields	Auction code, bid values, winning bid and number of bids per auction							
	All fields in the dataset	Common fields, PTE, difference Bid/PTE, location, Brazilian State and date	Common fields, PTE, difference Bid/PTE, location, type and size of bidding companies	Common fields, PTE, difference Bid/PTE, location and date	Common fields and consortium composition	Common fields, contract type and date	Common fields, bid value with and without inflation and date	Common fields only	
	Num. of variables	9	9	8	5	6	7	4	
	Screens	Coefficient of variation (CV), spread (SPD), percentage difference between the two lowest bids (DIFFP), relative distance (RD), skewness statistic (SKEW) and Kolmogorov-Smirnov test (KSTEST)							
Results. Best accuracy and top-performing algorithm	Setting 1 All fields from each dataset	85.2% Gradient Boosting	84.4% Extra Trees	94.7% Extra Trees	81.6% Bernoulli Naive Bayes	84.1% Ada Boost	84.1% Extra Trees	N/A	
	Setting 2 All fields from each dataset + screens	92.4% Gradient Boosting	87.4% Extra Trees	94.6% Gaussian Naive Bayes	91.4% Ada Boost	85.3% Extra Trees	84.8% Extra Trees	N/A	
	Setting 3 Common fields	87.9% Ada Boost	79.9% Random Forest	94.5% Extra Trees	82.0% Gaussian Naive Bayes	80.2% MLP	83.8% Extra Trees	82.0% Extra Trees	
	Setting 4 Common fields + screens	89.6% Extra Trees	86.8% Extra Trees	94.5% Extra Trees	91.4% Ada Boost	81.1% Extra Trees	83.7% Extra Trees	86.3% Extra Trees	
Average accuracy increase on including screens (for the four top-performing algorithms)	Best algorithms	Ensemble methods: Extra Trees, Random Forest, Ada Boost and Gradient Boosting							
	Setting 2 from 1	+6.0%	+2.3%	-0.5%	+12.1%	+0.1%	-0.1%	N/A	
	Setting 4 from 3	+1.6%	+2.5%	-0.9%	+14.1%	+1.9%	-0.7%	+1.1%	
Detection rates reported in the literature	Paper/s	[30,34]	[36]	[38]	[40]	[40]	[19]	N/A	
	Method	Probabilistic methods	Standard hierarchical clustering algorithm	ML methods: Random Forest & Ensemble Method	ML method: Random Forest	ML method: Random Forest	N/A	N/A	
	Accuracy	81% - 96%	N/A	88% - 93%	77% - 86%	61% - 84%	N/A	N/A	

Tabla 4.5: Resumen de los resultados de detección de colusión para los diferentes datasets (Brasil, Italia, Japón, Suiza-Ticino, Suiza-SG&GR, EE.UU. y en conjunto).

Los resultados para cada dataset se muestra en la Tabla 4.5. La última columna denominada *all datasets*, se refiere a que todos los datasets se han aglutinado en un único conjunto. Así se comprobará si el algoritmo es capaz de entrenarse con licitaciones de todo el mundo y detectar la colusión. Las filas que tienen el topic *Fields* se refieren a los campos que tiene cada dataset (no todos poseen los mismos campos) y que posteriormente se utilizarán como variables de entrada al algoritmo. En la siguiente fila *Results* se muestran el mejor porcentaje de acierto y el algoritmo que lo consiguió para los 4 settings. En las filas *Average accuracy* se muestra el incremento de la precisión media (calculada para los 4 mejores algoritmos) al añadir los screens a las variables de entrada, es decir, el incremento de la precisión del setting 2 respecto al 1 y del setting 4 respecto al 3. Por último, la fila *Detection rates reported in the literature* menciona la precisión y algoritmo obtenidos por otros artículos de la literatura, para que sirva como elemento de comparación del dataset utilizado.

En general viendo la Tabla 4.5, los mayores porcentajes de acierto se dan en el setting 2 y 4 (salvo para Japón y US que todos los settings están muy igualados). Esto significa que las variables de cribado (screens) ayudan a los algoritmos a aumentar la detección de colusión. En general, los algoritmos con mayor porcentaje de acierto son de tipo ensemble (Gradient Boosting, Ada Boost, Extra Trees y Random Forest). Para un mayor análisis de cada algoritmo, dataset, setting y métrica de error (precisión, falsos positivos, falsos negativos y precisión balanceada), consultar las Tablas 2 y 3 del artículo. En dichas tablas se observa que:

- La precisión es generalmente superior al 80%.
- Los falsos positivos (FP) y falsos negativos (FN) son generalmente inferiores al 10%.
- La precisión balanceada es generalmente superior al 70%.

La Figura 4.8 identifica otras tres métricas de error (precision, recall y F1 score) para los settings 3 y 4 del conjunto de datos denominado *all datasets*. Algunos algoritmos no aparecen en la gráfica por estar por debajo del 50% de precision y recall. Para que los resultados no sean sensibles a una sola iteración (recordemos, 80% de los datos para entrenamiento y 20% para test), se han realizado 500 iteraciones con cada algoritmo y, por tanto, hay 500 valores de precision y otros 500 de recall. Dichas métricas de error se indican con una cruz: el punto de corte es la mediana de la precision y recall y los puntos finales de la cruz son los valores mínimo y máximo de la precision y recall. Consecuentemente, los valores de precision, recall y F1 score permanecerán dentro del rectángulo formado por la cruz con un alto grado de confianza. Se observa como 3 de los ensemble methods (Ada Boost, Extra Trees, Random Forest) son los que mejor porcentaje obtienen (superiores al 80%) y, algo peor, el otro ensemble method (Gradient Boosting). El resto de algoritmos tienen unos porcentajes de precisión y recall muy inferiores.

En conclusión, los porcentajes de aciertos son notables y confirman que los algoritmos de ML se pueden aplicar de manera exitosa a la detección de colusión. En el material suplementario del artículo se adjunta para cada dataset sus histogramas, gráficas de error como el de la anterior Figura 4.8 y los boxplots de los screens. Reseñar que los datasets y el código programado también se han liberado públicamente como material suplementario, para que cualquier investigador pueda reproducir la experimentación o probar otros nuevos. Se ha tomado esta decisión por el hecho ya mencionado, la extrema dificultad en conseguir datasets que contengan licitaciones colusivas. Con esta decisión se quiere fomentar y facilitar el desarrollo de esta disciplina.

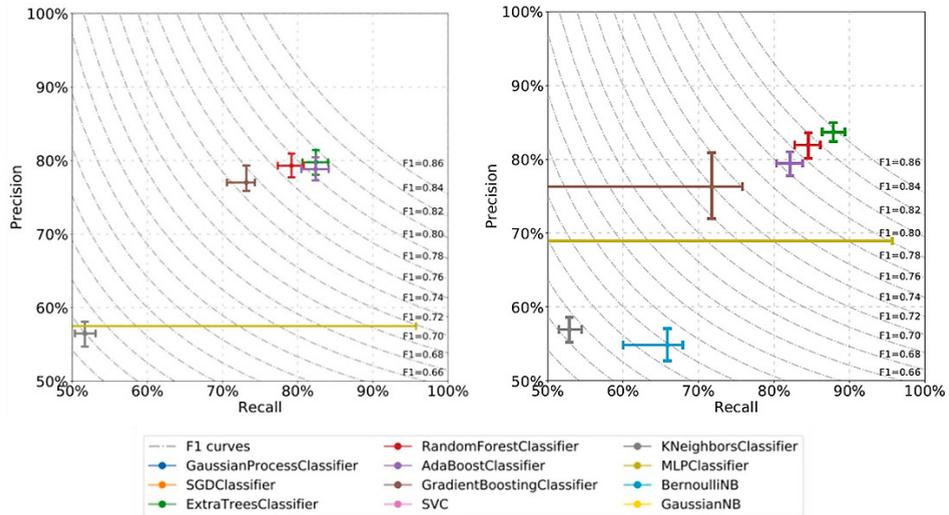


Figura 4.8: Gráficas del error (precision, recall y F1 score) para el dataset conjunto. A la izquierda para el setting 3 (campos comunes) y a la derecha para el setting 4 (campos comunes + screens).

4.6. APLICACIÓN INFORMÁTICA PARA DETECTAR LICITACIONES IRREGULARES

Finalmente, se comentará una investigación derivada de los conocimientos y experiencia acumulada gracias a la Tesis. Se ha desarrollado una aplicación que detecta pliegos técnicos irregulares y que, por tanto, es de sumo interés para los órganos públicos que velan por la contratación así como para la ciudadanía que quiere una contratación más justa y transparente. Aunque esta herramienta no ha derivado en la redacción de un artículo académico (por falta de tiempo), pone de manifiesto la utilidad de esta Tesis, su aplicación práctica y el gran interés por parte de los medios de comunicación y de la sociedad en general.

Esta herramienta fue desarrollada desde cero y totalmente ad hoc en el verano de 2021. Sin entrar en sus particularidades técnicas, se resume a alto nivel los pasos llevados a cabo:

1. La aplicación descargó masivamente documentos asociados a licitaciones publicadas de 2015 hasta 2021, gracias a los datos en abierto de PLACSP. Alrededor de un millón de documentos (pliegos técnicos, pliegos administrativos y otros documentos) fueron descargados.
2. De cada documento se leían sus propiedades (también llamados metadatos) que incluye el autor, la organización (empresa) y fecha de creación del documento. Estos metadatos pasan inadvertidos para personas poco

cuidadasas o inexpertas en informática, por no estar visibles en el propio documento digital. Nótese que este estudio tiene una importante limitación debido a que muchos de los documentos de PLACSP están escaneados (no es el documento digital original), perdiéndose los metadatos al escanearse.

3. Se compara la organización que aparece en los metadatos del pliego técnico con el nombre del adjudicatario (razón social) de la licitación²⁵. No sólo se compara con el adjudicatario sino también con el grupo empresarial al que pertenece. Dicha relación se establece gracias a la información empresarial obtenidas del Registro Mercantil (la otra fuente de información utilizada). Estas comparaciones se han realizado mediante un algoritmo de IA que es capaz de relacionar dos palabras, aunque no estén escritas exactamente igual.

No tiene sentido que un pliego técnico de una licitación con procedimiento abierto o abierto simplificado contenga el nombre de la empresa ganadora²⁶. Esto indicaría que el redactor de dicho pliego podría haber tenido alguna vinculación con el adjudicatario previamente a la publicación de la licitación. Que la relación entre el órgano de contratación y el ganador sea fraudulenta o no es algo que los metadatos no pueden demostrar. Lo que sí se puede afirmar es que el redactor del pliego utilizó como base un documento del adjudicatario. Por tanto, merecería la pena investigar si hubo vinculaciones irregulares entre ambos. Se detectaron varias licitaciones donde el nombre de la empresa ganadora figuraba en los metadatos del pliego y algunas de ellas fueron investigadas por el periódico *El País* que publicó una noticia²⁷ el 10 de diciembre de 2021 (léase en la Figura 4.9).

Si esta exitosa herramienta de detección ha sido desarrollada por un ciudadano con datos en abierto y escasos recursos, imaginemos los programas informáticos que podrían desarrollar los organismos que velan por la contratación. Organizaciones altamente especializadas que disponen de grandes recursos económicos, técnicos, humanos y que además tienen acceso a bases de datos privadas de las AA.PP.

²⁵. No se ha utilizado el autor que figura en las propiedades del documento porque hoy en día no se puede relacionar el nombre de una persona con la empresa donde trabaja. Si se podría automatizar con un programa informático que tenga acceso a las bases de datos de la Agencia Tributaria o la Seguridad Social que sí disponen de la relación persona-empresa.

²⁶. La LCSP en su artículo 115 (Consultas preliminares del mercado) establece el mecanismo de cómo los órganos de contratación pueden realizar consultas: “Los órganos de contratación podrán realizar estudios de mercado y dirigir consultas a los operadores económicos que estuvieran activos en el mismo con la finalidad de preparar correctamente la licitación”. Estas consultas deberán publicarse en el perfil del contratante.

²⁷. Accesible en <https://elpais.com/economia/2021-12-10/los-pliegos-de-tres-concursos-publicos-inclui-an-ganador-en-los-metadatos-antes-de-adjudicarse.html>

Los pliegos de tres concursos incluían el ganador en los datos digitales

Una investigación revela anomalías en al menos una convocatoria de RTVE, otra del Ministerio para la Transición Ecológica y una tercera de un consorcio catalán de hospitales

JOAQUÍN GIL, Madrid
Los pliegos de tres concursos públicos incluyeron en las propiedades del documento informático el nombre de la empresa ganadora. Así sucedió en Radiotelevisión Española (RTVE), en la Mancomunidad de los Canales del Taibilla, que depende del Ministerio para la Transición Ecológica, y en el Consorcio Sanitario del Maresme, una entidad controlada por la Generalitat catalana que gestiona una decena de centros de salud de Barcelona.

Los datos se desprenden del estudio de tres adjudicaciones por valor de 1,5 millones concedidas entre 2018 y 2020 y subidas a la Plataforma de Contratación del Sector Público por los organismos licitantes. Los casos analizados eran procedimientos abiertos, donde puede concurrir cualquier empresa, según establece la Ley de Contratos del Sector Público de 2017.

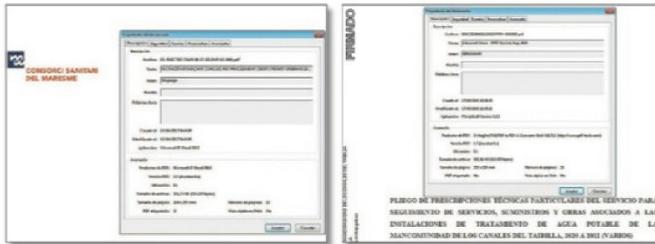
Una investigación de EL PAÍS a través del sistema de análisis masivo de expedientes desarrollado por el experto en Contratación Pública de la Universidad de Oviedo Manuel García Rodríguez y el científico de datos José Carlos Montes Luna ha hallado las coincidencias. Las propiedades de los archivos de los pliegos técnicos —el documento que recoge los requisitos que debe cumplir el contrato— revelan en unos casos que el nombre de la empresa ganadora aparecía en el campo "título" de las propiedades del PDF. Y en otros, que un equipo informático con el mismo nombre de la firma agraciada creó, alteró o guardó el archivo del pliego técnico meses antes de la presentación oficial de ofertas, la adjudicación o la formalización de los contratos. Estos son los casos analizados:

Mancomunidad de los Canales del Taibilla. Es un organismo adscrito al Ministerio de Transición Ecológica que abastece de agua potable a 2,4 millones de habitantes en Murcia, Alicante y Albacete. En el mes de junio de 2019 anunció una licitación para contratar a una firma de ingeniería para los años 2020 a 2022. El archivo del pliego técnico, un PDF de 35 páginas, contenía la palabra "Grusamar" en el campo autor. Fue guardado el 17 de mayo de 2019 a las 11.45. Cinco meses después, Grusamar Ingeniería y Consulting SAU, una de las dos firmas candidatas, obtiene esta adjudicación por 1.061.042 euros.

La Mancomunidad esgrimió que el vencedor presentó el pro-



El complejo Torrespaña de Radiotelevisión Española, en Madrid. / JAMES VILLANUEVA



Capturas de pantalla con las propiedades de los archivos informáticos que incluyen los pliegos de dos de los concursos.

yecto con la mejor relación calidad-precio. Sin embargo, la mención a Grusamar figuraba en los metadatos (huella digital) del expediente tres semanas antes de que la publicación oficial de la licitación se enviara al *Diario Oficial de la UE*, dos meses antes del fin del plazo de recepción de ofertas y tres de que se abriesen los proyectos de los aspirantes.

Un portavoz del Ministerio de Transición Ecológica justificó así la coincidencia: "Es posible que el técnico encargado de la tramitación del contrato utilizara como plantilla un pliego ya existente que podía arrastrar metadatos que le pasaron inadvertidos". Añadió que Grusamar trabaja para la Mancomunidad de los Canales del Taibilla "desde hace muchos años" como "apoyo en la parte administrativa" y precisó que el pliego fue elaborado por un técnico del organismo público. EL PAÍS ha intentado sin éxito recabar la versión de Grusamar.

RTVE. La dirección de compras de la corporación audiovisual

En dos de los casos, la adjudicataria de la licitación figuraba en el campo autor

Las firmas y los entes afectados niegan irregularidades en los procesos

anunció el 27 de mayo de 2020 la licitación de un contrato abierto "de transporte por satélite de la señal para Norteamérica y Centroamérica de RTVE". Y calculó un valor de 272.683 euros. El organismo público divulgó ese mismo día el pliego técnico, un documento de 34 páginas que recogía los requisitos que debía cumplir el ganador. El documento informático, que fue creado dos días antes del anuncio oficial de la licitación, contenía en sus propiedades los términos "Servicios Audiovisuales Overon, SL". Dos meses después, esta compañía de telecomunicaciones por satélite con 290 empleados ganaba una adjudicación de 246.996 euros para prestar el servicio. Fue la única empresa que acudió al concurso. Servicios Audiovisuales Overon SL es una firma del grupo Imagina Media Audiovisual SA, que está administrada por el empresario y productor de cine Jaime Roures, según el registro mercantil.

Un portavoz de RTVE asegura que la corporación desconoce a qué responde el hecho de que

el nombre de la empresa figure en las propiedades del archivo del pliego técnico. Y dice que la convocatoria se comunicó también a las compañías Eurovisión, Eutelsat, Eirona, Intelsat, Telefónica, además de a la ganadora. "Las empresas no conocían el pliego técnico", defiende el portavoz. Preguntado sobre la coincidencia, una representante de Mediapro, el grupo audiovisual de Roures, respondió: "No tenemos ni idea del PDF ni de los metadatos. Nosotros no hemos creado ese documento".

Consorcio Sanitario del Maresme. Se trata de un organismo que depende de la Consejería de Sanidad de la Generalitat catalana y que gestiona una decena de centros de salud, entre ellos el Hospital de Mataró. De sus servicios se benefician un total de 713.014 ciudadanos. La entidad sacó a concurso en 2017 un contrato bienal para adjudicar la explotación de máquinas expendedoras de bebidas en las instalaciones sanitarias. Las empresas aspirantes, en este caso, ofertaban cuánto estaban dispuestas a abonar al organismo público a cambio de colocar sus máquinas de venta de café o de agua.

El archivo del pliego técnico de la convocatoria se creó el 17 de octubre de 2017 —dos meses antes de la apertura oficial de plicas— e incluía en el campo autor de las propiedades la palabra "Gruparpa". Se trata —según el registro mercantil— del nombre de la matriz de la firma de vending (venta con máquinas expendedoras) Arbitrade SA, que medio año después ganaba una adjudicación de 387.200 euros para prestar este servicio y superaba a otras seis que aspiraban también a hacer negocio con la venta de café y botellas de agua en los hospitales.

El Consorcio Sanitario del Maresme niega que la entidad informara al ganador del pliego técnico antes de la convocatoria. "El contenido del concurso de 2018 se elaboró íntegramente desde el Consorcio, sin la contribución de ningún agente externo y siguiendo escrupulosamente lo que marca la ley en este sentido. Por lo tanto, no se informó al adjudicatario de su contenido en ningún momento hasta su publicación", responde un portavoz de este organismo por correo electrónico.

El vencedor del concurso de las máquinas expendedoras atribuye que su nombre aparezca como autor del pliego técnico a un error informático. "Las administraciones y los adjudicatarios intercambian documentos constantemente y a veces se produce la aparición del nombre de nuestra matriz en la autoría del archivo es una pura anécdota. Las empresas facilitamos información en archivos informáticos constantemente sobre máquinas, productos, listas, precios, características, o cualquier otro detalle que precise la administración", añade un representante de la empresa Gruparpa.

Figura 4.9: Aplicación informática para detectar licitaciones irregulares recogida en la noticia de El País publicada el 10/12/2021.

CAPÍTULO QUINTO. DISCUSIÓN DE LOS RESULTADOS

Este capítulo pretende hacer una discusión de los artículos en su conjunto, subrayando los elementos comunes y transversales de las técnicas de ML aplicadas al campo de la contratación pública. Para un mayor detalle técnico, léase el apartado de discusión de resultados de cada artículo. Además, se va a realizar ciertos análisis desde la perspectiva económica, estudios que apenas se ha hecho en los artículos y que, por tanto, los complementa. Así, se apreciarán los beneficios de aplicar la ciencia de datos a los contratos públicos. Es decir, los académicos de las ciencias sociales y económicas deben incorporar las herramientas analíticas, cuantitativas, para profundizar en sus investigaciones ligadas a la contratación.

Los artículos de la Tesis ponen de manifiesto la capacidad que tiene el Machine Learning (ML) para resolver problemas característicos de la contratación. En particular, se han tratado tres tipos de cuestiones: estimar el precio de adjudicación, buscar empresas que puedan llevar a cabo una licitación y detectar prácticas colusorias (anticompetitivas) en las licitaciones. Se podrían haber escogido otros problemas porque hay muchas preguntas a resolver en este ámbito. La aplicación de ML a este campo es algo todavía novel, incipiente, que están descubriendo las AA.PP y el sector privado. Esta Tesis pretende servir de ejemplo y abrir las puertas a futuros investigadores y estudios. Todo lo que rodea a la contratación tiene un gran impacto económico y mucha relevancia dentro de la sociedad.

Se va a tratar dos cuestiones capitales para la contratación y que, consecuentemente, tienen especial interés para las AA.PP. y relevancia para la sociedad:

1. **Mejorar la asignación de los recursos económicos** mediante un cálculo del presupuesto global de contratación más realista. Es decir, que las AA.PP. sean capaces de hacer una mejor planificación económica de sus compras públicas utilizando precios de mercados.

2. **La concurrencia en los contratos y sus efectos económicos.** Es decir, analizar cuantitativamente el impacto que tiene la concurrencia (el número de ofertantes) en los importes de las licitaciones. Por qué el aumento de la competencia tiene un efecto beneficioso.

Comencemos por la cuestión del primer punto. Un presupuesto de contratación global lo más realista posible es que la AA.PP. correspondiente sea capaz de presupuestar todas las licitaciones lo más aproximado al precio final de adjudicación, de mercado. Para ello, se necesitan herramientas como el estimador del importe de adjudicación, objeto de estudio en los artículos [10] y [11] de la Tesis. Por tanto, el elemento fundamental de análisis es la diferencia económica entre el importe de licitación (presupuesto) y el importe de adjudicación. A esto se le llama baja económica.

Se muestra en la Figura 5.1 un diagrama de cajas (boxplot; para más detalle consultar la Figura 3.5) de la baja económica (en azul), es decir, el error porcentual absoluto (APE) entre el importe de adjudicación y de licitación. Además, en gris se muestra la baja económica entre el importe de adjudicación y la predicción del estimador. Se muestran desagregadas por CPV (Common Procurement Vocabulary), es un código de 8 dígitos que clasifica la naturaleza de las licitaciones (el tipo de trabajo o servicio a realizar). Se han utilizado solamente la agregación de los dos primeros dígitos del CPV para tener un grupo de clasificación manejable y así poder representar gráficamente. Por ejemplo, el CPV 45 es “trabajos de construcción” y el 71 es “Servicios de arquitectura, construcción, ingeniería e inspección”²⁸. Finalmente, se muestra el error porcentual absoluto medio (MAPE) y el error porcentual absoluto mediano (MdAPE). La Figura ha sido extraída del primer artículo [10], por tanto se basa en 58.337 licitaciones de España (para más detalle consultar dicho artículo).

Se observa en la ya mencionada Figura 5.1 que el rango intercuartil (IQR) es generalmente más pequeño si se utiliza el estimador del importe de adjudicación (en gris). Esto se confirma al ver que el MAPE y MdAPE suelen ser también inferiores (cuantificado en la Tabla 4.1). El MdAPE se reduce un -2,58% [10], esto significa que las AA.PP. españolas podrían haber reducido unos 811 millones de euros en sus presupuestos iniciales, pudiendo destinar ese importe a otras partidas. Recordar que el primer artículo [10] utiliza el algoritmo Random Forest, pero éste fue posteriormente superado por la ANN y la regresión isotónica en el segundo artículo [11]. Por tanto, el IQR en gris sería incluso mejor (más pequeño) que el IQR en azul si se empleasen los algoritmos del segundo artículo [11].

²⁸ La descripción completa de códigos CPV está en el [Reglamento \(CE\) núm. 213/2008 de la CE](#), de 28 de noviembre de 2007, por el que se aprueba el vocabulario común de contratos públicos.

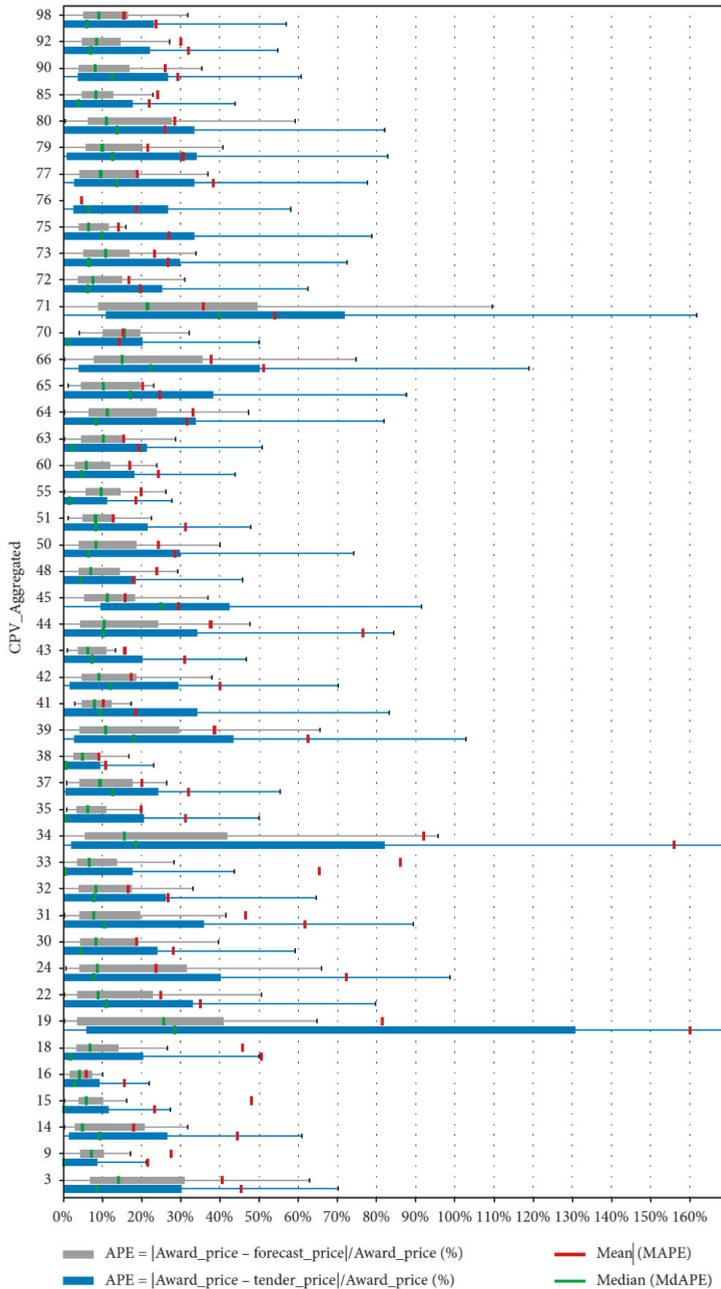


Figura 5.1: Diagrama de cajas (boxplot) del error porcentual absoluto (APE) entre el importe de adjudicación y la predicción (gris) y el APE entre el importe de adjudicación y licitación (azul). Agrupados por CPV para los datos de España. Fuente: elaboración propia en [10].

Un estimador del importe de adjudicación, además de la utilidad para las AA.PP., es una herramienta útil para los ofertantes (empresas), ayudándolas en su toma de decisiones y tratando de reducir la incertidumbre económica. Sin embargo, es irreal pensar que se pueda diseñar un estimador del importe de adjudicación perfecto, sin error. El mercado es un lugar abierto, libre, competitivo, sin un intercambio de información perfecta, donde los operadores económicos llevan a cabo sus acciones y decisiones. Además, el importe de adjudicación no es el finalmente pagado a los adjudicatarios. Después de la firma del contrato puede haber modificaciones durante la ejecución del contrato, repercutiendo en un cambio del importe (sobrecostes). Actualmente, en España no se publican los datos en abierto de las modificaciones del contrato y, por tanto, no se puede hacer un estudio masivo de licitaciones aplicando la ciencia de datos. Es una pena que las modificaciones no se publiquen de manera masiva y estructurada porque de su estudio y análisis se derivaría un mejor gasto público al caracterizar dichos contratos públicos.

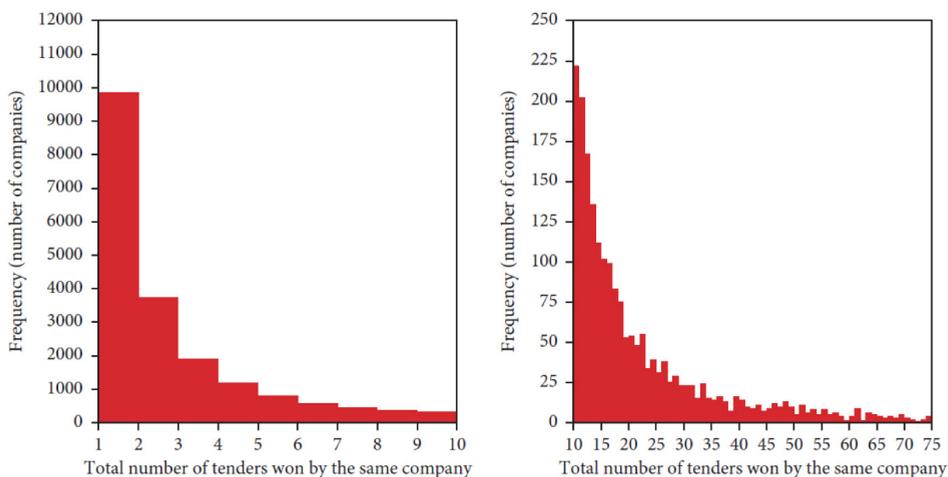


Figura 5.2: Histograma que representa el número de empresas (eje y) que han ganado el mismo número de licitaciones (eje x). Se ha dividido el gráfico en dos para una mejor visualización. Fuente: elaboración propia en [12].

La otra cuestión que se va a estudiar a través de los artículos de la Tesis versa sobre la concurrencia en los contratos y sus efectos económicos en España (extrapolable a otros países). En la Figura 5.2 se representa en un gráfico de barras cuántas empresas han ganado el mismo número de licitaciones [12]. Por ejemplo, el mayor número de empresas (casi 10 mil) han ganado una única licitación. Cerca de 4 mil empresas han ganado 2 licitaciones y así sucesivamente. Esto indica que en España hay una gran fragmentación de proveedores para las AA.PP., es decir, hay una considerable rotación de proveedores. En general, un mismo proveedor no es contratado por

las AA.PP., no generándose sinergias por la recurrencia. Esto puede ser a causa de varios factores: los contratos públicos no son atractivos para las empresas, hay una alta competencia en el mercado (posteriormente se demostrará que esta afirmación es falsa), los adjudicatarios trabajan una vez con la AA.PP. y luego no repiten por las razones que sean (mala experiencia en el negocio, pocos beneficios, exceso de regulación, etc.), los adjudicatarios están débiles empresarialmente (no pueden competir en el mercado privado) y recurren como última instancia a llevar a cabo una licitación pero acaban quebrando, etc. Se debería hacer un análisis económico-financiero de los adjudicatarios y una encuesta a las AA.PP. y empresarios para conocer las verdaderas causas.

En la Tabla 5.1 se dividen las licitaciones (analizadas en [10]) en 4 grupos:

Description	Groups by competitiveness			
	No competitiveness	Low Received offers (2-4)	Medium Received offers (5-10)	High Received offers >10
Total number of tenders in the dataset	18,790	22,714	11,553	5,271
Total number of tendering organisations	1,956	2,553	2,135	1,053
Total number of winning/award companies	7,550	9,555	5,222	2,402
Mean received offers by tender	1.0	2.80	6.73	20.01
Mean duration of tender's works	401.07 days	396.65 days	370.95 days	277.50 days
Mean tender price	€354,882.49	€388,526.27	€785,455.49	€1,301,031.70
Median tender price	€60,500.00	€75,000.00	€121,000.00	€254,376.00
Mean award price	€341,874.79	€323,611.87	€460,548.68	€836,188.79
Median award price	€58,984.50	€64,833.00	€90,689.00	€174,986.00
Median absolute error (MdAE)	€93.50	€7,661.50	€22,854.00	€76,420.00
Median absolute percentage error (MdAPE)	0.12%	13.39%	29.63%	45.94%
Mean absolute error (MAE)	€13,966.65	€68,244.60	€326,698.33	€464,907.75
Mean absolute percentage error (MAPE)	10.02%	25.65%	54.48%	77.98%

Tabla 5.1: Diferencia entre el importe de licitación y adjudicación (baja económica) para diferentes grupos de ofertas recibidas. Fuente: elaboración propia en [10].

1. Licitaciones que sólo han recibido una oferta (no hay concurrencia).
2. Licitaciones que han recibido entre 2 y 4 ofertas (baja concurrencia).
3. Licitaciones que han recibido entre 5 y 10 (media concurrencia).
4. Licitaciones que han recibido más de 10 (alta concurrencia).

De esta manera, se puede evaluar fácilmente la competencia que ha habido en la contratación. Se aprecia que las licitaciones tienen un importe mediano (tanto el importe de licitación como el importe de adjudicación) creciente, es decir, cuanto mayor es el grupo de concurrencia, mayor es el importe mediano. Lo mismo sucede con la baja económica media (el MAPE): aumenta cuanto mayor es el grupo de concurrencia. Para el grupo 1 (sin concurrencia) la baja es el 10,02%, para el grupo 2 es 25,65%, para el grupo 3 es 54,48% y para el grupo 4 es 77,98%. Por tanto, se aprecia claramente el efecto de que a una mayor concurrencia, competitividad, las licitaciones se logran adjudicar con una baja mucho mayor, consiguiendo las AA.PP. un importante ahorro presupuestario.

Continuando con el mismo análisis, en la Figura 5.3 aparecen dos gráficas que representan el número de ofertas (concurrencia) y la baja económica media. En la

gráfica superior (elaboración propia en [12]) se observa como la baja crece hasta estabilizarse entre el 30% y 40%, aproximadamente. Esta estabilización ocurre a partir de las 20 ofertas por licitación. La gráfica inferior es de la Oficina Independiente de Regulación y Supervisión de la Contratación (OIReScon) publicada en su “Informe Anual de Supervisión de la contratación pública en España” de 2022 [21] y, por tanto, está elaborada con una mayor muestra de licitaciones. Ambas gráficas tienen una tendencia similar y corroboran que a una mayor concurrencia, una mayor baja. Dicha baja se estabiliza a partir de un número de ofertas, 20 aproximadamente.

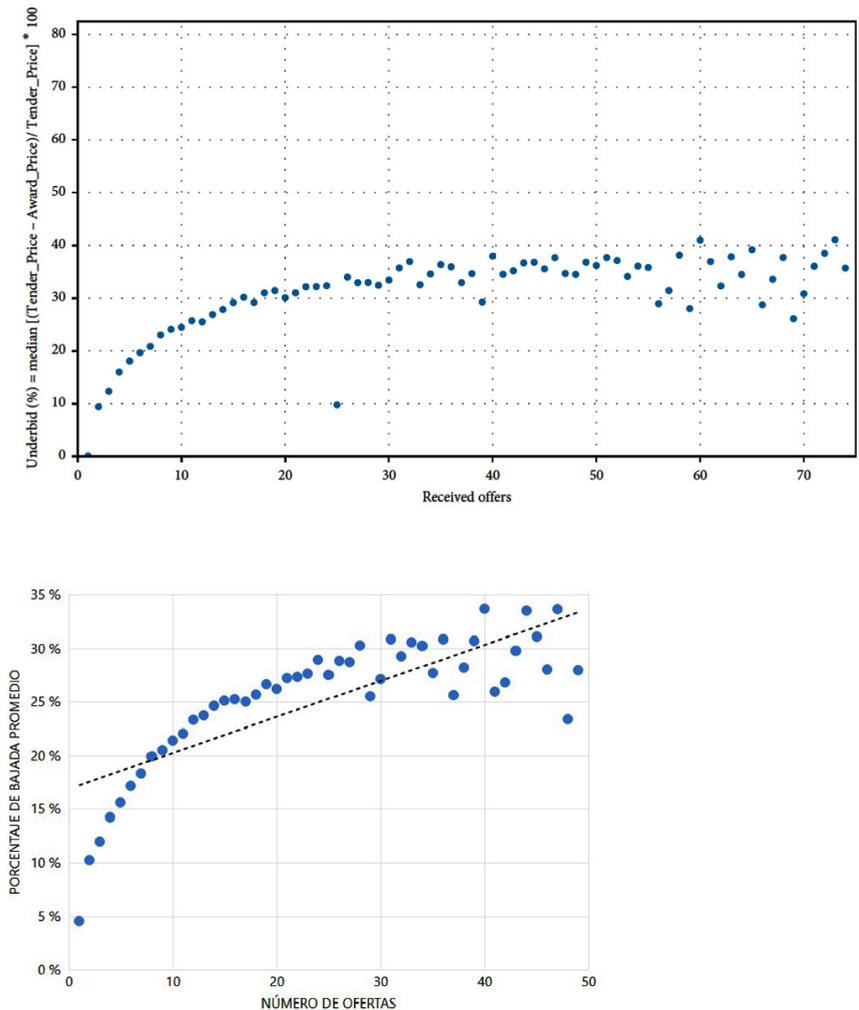


Figura 5.3: Diferencia porcentual entre el importe de licitación y adjudicación (baja económica) (eje y) según el número de ofertas recibidas en la licitación (eje x). La gráfica superior es la baja mediana (elaboración propia en [12]) y la gráfica inferior es la baja media (elaborada por la OIReScon en [21]).

Por tanto, el fenómeno económico de que en un mercado libre con una gran competencia produce unos bienes y servicios de mejor calidad a menores precios, se confirma también en el mercado de la contratación pública. Esta relación se ha demostrado cuantitativamente en artículos de otros investigadores, por ejemplo, para Italia [110] o República Checa [111, 112]. Y se menciona expresamente “*de mejor calidad*” porque las licitaciones no sólo tienen como criterio de adjudicación el precio, sino también valoran la calidad a través de los criterios de adjudicación técnicos y subjetivos. Sin embargo, aproximadamente el 32% de licitaciones tienen solamente una única oferta, como se ha observado en la Tabla 5.1 (primera fila, expresado en porcentaje). Este porcentaje también se corrobora en la Figura 5.4, elaborada por la Comisión Europea (CE)²⁹ con datos oficiales de TED, donde España tiene un 28% y ha aumentado desde el 2017 al 2020. Se deberían de implementar políticas de contratación que fomenten la participación de empresas porque si no se lastra la contratación. Por esa razón, se diseñó el buscador/recomendador de licitadores [12], herramienta que sirve para fomentar la competitividad. Además, en los escenarios de alta participación es más difícil que se produzca corrupción o colusión por ser un entorno donde, al operar muchas empresas, no se puede manipular la licitación fácilmente.

Se podrían haber estudiado más aspectos económicos analizando los datos de contratación pero hay que poner un límite y, además, queda fuera del alcance de la Tesis. Se han descrito los anteriores ejemplos para resaltar cómo se pueden hacer análisis cuantitativos en la contratación y los beneficios que aportan.

²⁹ Accesible en la web de la CE [Single market scoreboard-public procurement](#)

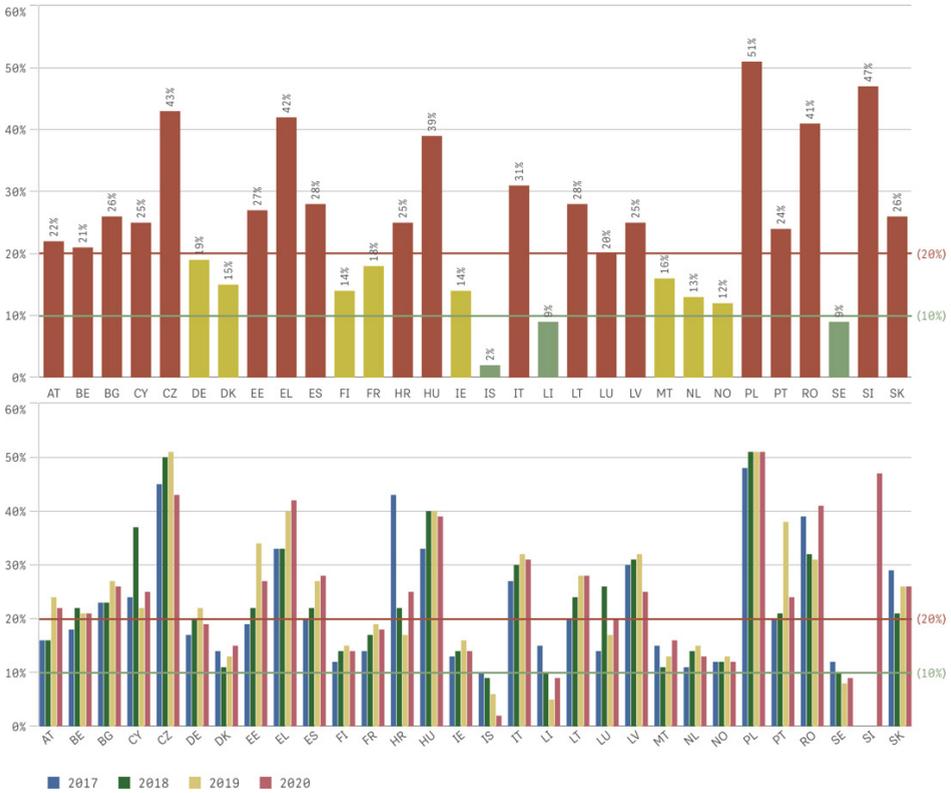


Figura 5.4: Porcentaje de licitaciones que han tenido una única oferta, dividido por países y años. Fuente: CE.

CAPÍTULO SEXTO. CONCLUSIONES

6.1. CONCLUSIONES

La contratación pública es un campo de gran importancia por representar un porcentaje significativo del gasto sobre el PIB de los Estados, un 16% según algunas estimaciones oficiales [2] pero puede ser incluso más. La contratación, vista solamente como las leyes de contratos del Sector Público, ha sido estudiada por juristas, expertos en Derecho Administrativo. Sin embargo, es un campo muy poco estudiado por los investigadores del ámbito científico-tecnológico, presumiblemente porque hasta hace pocos años no se disponían de datos de licitaciones, tanto a nivel español como de la UE u otros países.

Esta Tesis es pionera en la investigación científica-tecnológica sobre la contratación pública, al abordarse desde la ciencia de datos y desarrollar nuevas herramientas gracias a las tecnologías digitales como el procesamiento masivo de datos y el Machine Learning (ML). No se conocen Tesis de similares características en España ni tampoco se han encontrado fuera. Esto queda de manifiesto en la revisión de la literatura académica mencionada en los 4 artículos de la Tesis, que es muy escasa a nivel internacional y prácticamente inexistente a nivel nacional. Tampoco abundan informes técnicos que utilicen dichas tecnologías, redactados por expertos técnicos o por las propias Administraciones Públicas (AA.PP.). No obstante, en los últimos años han surgido iniciativas y proyectos innovadores dentro de las AA.PP., desarrolladas en su mayoría por el sector privado, que aplican las nuevas tecnologías digitales para crear funcionalidades innovadoras que mejoren la gestión y supervisión de la contratación.

A lo largo de los capítulos de la Tesis, se ha introducido la importancia de la contratación, la ciencia de datos y brevemente sus evoluciones históricas. Se han

descrito los retos actuales de la contratación, qué organismos públicos de España están ligados a ella, la transparencia y datos en abierto de España y las iniciativas de la Comisión Europea (CE) asociadas a los datos de contratación. Se han enumerado las causas de por qué la contratación es ahora un área digital y cuantitativa, su legislación, tecnologías aplicables a la contratación (Big Data, Data Analytics, IA, NLP, Process mining y RPA) y casos de uso que ponen en práctica dichas tecnologías. Se ha descrito en detalle la Plataforma de Contratación del Sector Público (PLACSP) desde la perspectiva de los datos en abierto: órganos de contratación que publica en PLACSP, origen, formato y calidad de los datos y el programa OpenPLACSP. Se ha recopilado literatura académica asociada a la contratación sobre 4 temáticas: datos en abierto y calidad del dato, innovación y gestión en la contratación, forecasting en la contratación y colusión y corrupción. Además, se formulan varias métricas de error típicas para resolver los problemas tratados en la Tesis (regresión y clasificación) y así poder evaluar y comparar los algoritmos de ML.

Se han descrito los 4 artículos que forman la Tesis de manera asequible a un lector no técnico. De esta manera se abre la Tesis a investigadores de otros ámbitos, como el jurídico o económico. El lector técnico puede profundizar más leyendo en detalle cada artículo, para analizar cómo se resuelve el problema y mediante qué técnicas de ML. Los artículos se titulan en español: “*Licitaciones en España: regulación, análisis de datos y estimador del importe de adjudicación usando ML*”, “*Estimador del importe de adjudicación de licitaciones usando ML: caso de estudio con licitaciones de España*”, “*Recomendador de licitadores usando ML: análisis de datos, algoritmo y caso de estudio con licitaciones de España*” y “*Detección de colusión en licitaciones aplicando algoritmos de ML*”. Además, se ha descrito la investigación titulada “*Aplicación informática para detectar licitaciones irregulares*”, que de momento no ha dado lugar a un artículo académico pero sí ha detectado licitaciones españolas donde ha podido existir actuaciones ilegales, fraudulentas.

Los artículos presentados son ejemplos útiles y prácticos de cómo la ciencia de datos y el ML ayudan a los diferentes actores de la contratación (AA.PP, empresas, investigadores, etc.) y, en última instancia, a la sociedad que sufraga con sus impuestos la compra pública y los padece si son de mala calidad. En los artículos se emplearon dos tipos de fuentes de datos: datos de contratación (PLACSP, TED y datasets de colusión) y las cuentas anuales de empresas (alojadas en el Registro Mercantil de España). Por tanto, los datos en abierto son fundamentales para poder desarrollar este tipo de investigación, realizada por una persona externa a las AA.PP. y que sólo dispone de información pública. La transparencia en la contratación y las leyes que obligan a publicar los datos en formatos abiertos y reutilizables juegan un papel fundamental en este sentido.

Además de la discusión de resultados en los propios artículos, principalmente sobre las técnicas de ML aplicadas y los porcentajes de acierto, en esta memoria de la Tesis se

ha incorporado una discusión desde un punto de vista económico. Se han tratado dos cuestiones relevantes. Por un lado, mejorar la asignación de los recursos económicos mediante un cálculo del presupuesto global de contratación más realista. Por otro lado, cómo el aumento de la concurrencia en los contratos tiene un beneficio económico positivo, es decir, cuantificar el ahorro obtenido debido a una mayor competitividad. Sin embargo, también se cuantifica el bajo número de ofertantes en las licitaciones, que impide la competencia efectiva en las licitaciones.

Las conclusiones particulares de cada investigación se detallan en dichos artículos (influencias de las variables, comparativa detallada de los modelos, etc.), no se ha hecho una mención explícita y repetitiva en esta memoria.

Finalmente, subrayar que la Tesis pretende servir de puente entre la contratación pública y la ciencia de datos, dos campos bastante distantes. El primero liderado por investigadores de las ciencias sociales (juristas, economistas, politólogos, etc.) y el segundo por investigadores del ámbito científico-tecnológico (ingenieros, matemáticos, informáticos, etc.). Por tanto, a esta Tesis se le denominaría de frontera, porque que ambos grupos pueden comprenderla y asimilarla para sus propias investigaciones. Se espera que sirva de ejemplo para futuros proyectos en colaboración, transversales y multidisciplinares.

6.2. LÍNEAS DE INVESTIGACIÓN FUTURAS

Como se ha mencionado en el apartado anterior, la investigación pretende servir de ejemplo para que otros investigadores del ámbito científico-tecnológico generen nuevos estudios, debido a que hay una gran carencia. Además, por la propia naturaleza de la contratación pública, las investigaciones tendrán un impacto significativo en las AA.PP., el mercado y la sociedad. Por tanto, hay un futuro prometedor y más si se tiene en cuenta que las AA.PP. están en un periodo de profunda digitalización, es decir, están aumentando la cantidad y calidad de los datos asociados a sus procesos administrativos y, particularmente, los de contratación.

Las líneas de investigación futuras que se deberían de llevar a cabo en España y en los países extranjeros pasan por seguir profundizando en el análisis de datos de contratación, tanto en la fase de licitación (pre-award) como en la fase de adjudicación y ejecución del contrato (post-award). Por ejemplo, analizar las desviaciones en plazos, costes y calidad de la ejecución de los contratos con respecto al proyecto inicial (presupuesto económico estimado y planificación de los trabajos). ¿Existe una relación entre proyectos iniciales mal realizados y, como consecuencia, se producen sobrecostes u otras desviaciones en la ejecución del proyecto?

A continuación, se enumeran acciones que se deberían realizar para promover y facilitar investigaciones futuras:

- Mayor transparencia en la contratación, abarcando todo el ciclo de compra. Es decir, que se publiquen en formato de datos en abierto (públicos y reutilizables) las modificaciones de los contratos y las facturas de los trabajos realizados. Con esta información se llevaría a una nueva dimensión la investigación relacionada con la contratación.
- Aumentar los campos disponibles de contratación. Por ejemplo, que cuando se adjudique una licitación se publique, de manera estructurada, la identidad de los ofertantes y sus respectivas ofertas económicas. Esto no se hace en la actualidad y es una información valiosa. Así se podrán elaborar estudios económicos sobre la competencia o desarrollar herramientas de detección de colusión más potentes aplicando ML.
- Mejorar la calidad del dato de la contratación. No sólo sucede en este campo, es un problema generalizado a las AA.PP. españolas porque actualmente no hay una cultura del dato. Faltan profesionales, aplicaciones y normativas que estandaricen cómo estructurar, almacenar y gestionar los datos.
- Las fuentes de datos públicas deben ser también gratuitas. Por ejemplo, para realizar estudios de contratación es fundamental conocer a los adjudicatarios, caracterizar a las empresas. Estos datos provienen del Registro Mercantil y, en la actualidad, son públicos pero de pago. De poco sirve que la ley disponga que los datos sean públicos si tienen un coste.
- Aumentar la interoperabilidad de los sistemas informáticos de las AA.PP. Así se conseguiría realizar estudios más elaborados y complejos, de más valor añadido, al incorporar fuentes de datos de distintas organizaciones o naturaleza.
- Colaboración entre los expertos de las ciencias sociales y del científico-tecnológico para abordar los retos de la contratación de una manera integral. Es decir, investigaciones realizadas por equipos multidisciplinares que trabajen de manera coordinada y conjuntamente.

BIBLIOGRAFÍA

- [1] F. Curtis y P. Maines, "Closed competitive bidding," *Omega*, vol. 1, n.º 5, págs. 613-619, oct. de 1973, issn: 03050483. doi: [10.1016/0305-0483\(73\)90049-2](https://doi.org/10.1016/0305-0483(73)90049-2).
- [2] "Public procurement contracts," *European Parliament, Fact Sheets on the European Union*, 2021. dirección: <https://www.europarl.europa.eu/factsheets/en/sheet/34/public-procurement-contracts>.
- [3] J. VanderPlas, *Python Data Science Handbook*, 1.ª ed. O'Really, 2017, pág. 548, isbn: 978-1-491-91205-8.
- [4] A. S. Patrucco, A. Moretto, S. Ronchi y D. Luzzini, "Organisational choices in public procurement: what can public management learn from the private sector?" *Local Government Studies*, vol. 45, n.º 6, págs. 977-1000, nov. de 2019, issn: 0300-3930. doi: [10.1080/03003930.2019.1608827](https://doi.org/10.1080/03003930.2019.1608827).
- [5] J. Miranzo Díaz, "Inteligencia Artificial Y Contratación Pública," en *Administración Electrónica, Transparencia y Contratación Pública*, Madrid: Iustel, 2020, págs. 105-142, isbn: 978-84-9890-385-0. dirección: <https://ssrn.com/abstract=3647414>.
- [6] L. Cotino Hueso y A. Todolí Signes, *Explotación y regulación del uso del big data e inteligencia artificial para los servicios públicos y la ciudad inteligente*. Valencia: Tirant lo Blanch, 2022, pág. 348, isbn: 9788411133159.
- [7] J. W. Tukey, "The Future of Data Analysis," *The Annals of Mathematical Statistics*, vol. 33, n.º 1, págs. 1-67, mar. de 1962, issn: 0003-4851. doi: [10.1214/aoms/1177704711](https://doi.org/10.1214/aoms/1177704711).

- [8] P. Naur, *Concise survey of computer methods*. Petrocelli Books, 1974, pág. 397, isbn: 9780884053149.
- [9] P. del consejo de Ministros, “Real Decreto Ley de Contratos sobre Servicios Públicos,” *Gaceta de Madrid*, vol. 6460, págs. 1-2, 1852. dirección: <https://www.boe.es/datos/pdfs/BOE/1852/6460/A00001-00002.pdf>.
- [10] M. J. García Rodríguez, V. Rodríguez Montequín, F. Ortega Fernández y J. M. Villanueva Balsera, “Public Procurement Announcements in Spain: Regulations, Data Analysis, and Award Price Estimator Using Machine Learning,” *Complexity*, vol. 2019, n.º v, 2019, issn: 10990526. doi: [10.1155/2019/2360610](https://doi.org/10.1155/2019/2360610).
- [11] M. J. García Rodríguez, V. Rodríguez Montequín, A. Aranguen Ubierna, R. Santana Hermida, B. Sierra Araujo y A. Zelaia Jauregi, “Award Price Estimator for Public Procurement Auctions Using Machine Learning Algorithms: Case Study with Tenders from Spain,” *Studies in Informatics and Control*, vol. 30, n.º 4, págs. 67-76, dic. de 2021, issn: 12201766. doi: [10.24846/v30i4y202106](https://doi.org/10.24846/v30i4y202106).
- [12] M. J. García Rodríguez, V. Rodríguez Montequín, F. Ortega Fernández y J. M. Villanueva Balsera, “Bidders Recommender for Public Procurement Auctions Using Machine Learning: Data Analysis, Algorithm, and Case Study with Tenders from Spain,” *Complexity*, vol. 2020, T. C. Silva, ed., págs. 1-20, nov. de 2020, issn: 1099-0526. doi: [10.1155/2020/8858258](https://doi.org/10.1155/2020/8858258).
- [13] M. J. García Rodríguez, V. Rodríguez-Montequín, P. Ballesteros-Pérez, P. E. Love y R. Signor, “Collusion detection in public procurement auctions with machine learning algorithms,” *Automation in Construction*, vol. 133, pág. 104 047, ene. de 2022, issn: 09265805. doi: [10.1016/j.autcon.2021.104047](https://doi.org/10.1016/j.autcon.2021.104047).
- [14] M. J. García Rodríguez, V. R. Montequín, F. O. Fernández y J. V. Balsera, “Spanish Public Procurement: Legislation, open data source and extracting valuable information of procurement announcements,” *Procedia Computer Science*, vol. 164, págs. 441-448, 2019, issn: 18770509. doi: [10.1016/j.procs.2019.12.204](https://doi.org/10.1016/j.procs.2019.12.204).
- [15] D. Arosa Otero, J. C. Arvelo Flores, M. Cano Rodríguez, A. Colomer Pedrosa y M. J. García Rodríguez, “La contratación pública en España: fuentes de datos, normativa y aplicaciones tecnológicas,” *Revista de la Escuela Jacobea de Posgrado*, vol. 21, págs. 87-112, 2021. dirección: <https://www.jacobeas.edu.mx/revista/numero21.php>.
- [16] M. J. García Rodríguez, “Tecnologías digitales para el control de la contratación pública,” en *Auditoria pública*, ASOCEX, ed., vol. 79, Sevilla, 2022, isbn: 1136-517 X. dirección: <http://auditoriapublica.com>.
- [17] C. Ramiró, *Inteligencia artificial y Administración pública: Robots y humanos compartiendo el servicio público*. Catarata, 2019, pág. 176, isbn: 978-84-9097-590-9.

- [18] P. Valcárcel Fernández, “Strategic and Smart Public Procurement. Big Data, Open Data, public procurement 4.0 ¿Una nueva era?” En *X Seminario de Contratación Pública. Nuevos retos de la contratación pública: discusión práctica*, 2019.
- [19] C. Europea, “Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones: Conseguir que la contratación pública funcione en Europa y para Europa,” inf. téc., 2017, pág. 17. dirección: <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=COM:2017:572:FIN>.
- [20] C. Europea, “Aplicación y mejores prácticas de las políticas nacionales de contratación pública en el mercado interior,” Comisión Europea, Bruselas, inf. téc., 2021, pág. 11. dirección: <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:52021DC0245>.
- [21] OIReScon, “Informe Anual de Supervisión de la contratación pública en España,” Oficina Independiente de Regulación y Supervisión de la Contratación (OIReScon), inf. téc., 2021. dirección: <https://www.hacienda.gob.es/RSC/OIReScon/informe-anual-supervision-2021/ias-2021.pdf>.
- [22] M. Mendes y M. Fazekas, “DIGIWHIST Recommendations for the Implementation of Open Public Procurement Data An Implementer’s Guide,” págs. 1-16, 2018. dirección: <https://opentender.eu/blog/2017-03-recommendations-for-implementation/>.
- [23] N. Modrušan, L. Mršić y K. Rabuzin, “Intelligent Public Procurement Monitoring System Powered by Text Mining and Balanced Indicators,” en 2021, págs. 115-133. doi: [10.1007/978-3-030-83014-4_6](https://doi.org/10.1007/978-3-030-83014-4_6).
- [24] OECD, *Public Procurement for Innovation* (OECD Public Governance Reviews). OECD, jun. de 2017, isbn: 9789264265813. doi: [10.1787/9789264265820-en](https://doi.org/10.1787/9789264265820-en).
- [25] C. Bovis, “The priorities of EU public procurement regulation,” *ERA Forum*, vol. 21, n.º 2, págs. 283-297, oct. de 2020, issn: 1612-3093. doi: [10.1007/s12027-020-00608-8](https://doi.org/10.1007/s12027-020-00608-8).
- [26] I. González Ríos, “La transparencia como principio vertebrador de la contratación pública: significado y problemas de articulación normativa,” *Revista de Estudios de la Administración Local y Autonómica*, págs. 6-25, oct. de 2019, issn: 1989-8975. doi: [10.24965/reala.i12.10714](https://doi.org/10.24965/reala.i12.10714).
- [27] M. Bauhr, Á. Czibik, J. Fine Licht y M. Fazekas, “Lights on the shadows of public procurement: Transparency as an antidote to corruption,” *Governance*, vol. 33, n.º 3, págs. 495-523, jul. de 2020, issn: 0952-1895. doi: [10.1111/gove.12432](https://doi.org/10.1111/gove.12432).

- [28] M. Fernández Salmerón y R. Martínez Gutiérrez, *Transparencia, innovación y buen gobierno en la contratación pública*. Valencia: Tirant lo Blanch, 2019, pág. 492, isbn: 9788491905943.
- [29] P. Valcárcel Fernández, “Transparency in public procurement in the Spanish legal system,” en *Transparency in EU Procurements. Disclosure rules within public procurement procedures and during contract period*, 4, K.-M. Halonen, R. Caranta y A. Sánchez Graells, eds., Edward Elgar Publishing, 2019, págs. 272-295, isbn: ISBN 978-1-78897-566-7. doi: [10.4337/9781788975674.00019](https://doi.org/10.4337/9781788975674.00019).
- [30] A. Cerrillo Martínez, “Contratación electrónica y transparencia: fundamentos necesarios de la contratación abierta,” *Cuadernos de derecho local*, págs. 121-149, 2018, issn: 1696-0955.
- [31] J. Valero Torrijos, “Transparencia, acceso y reutilización de la información del sector público,” en *Transparencia y acceso a la información pública: de la teoría a la práctica*, Iustel, ed., 2019, págs. 225-250, isbn: 978-84-9890-373-7.
- [32] P. Valcárcel Fernández, “Tres dimensiones de la transparencia en la contratación pública. Rendición de cuentas, respeto de los derechos de operadores económicos y mejora global de la gestión de este sector a través del big data,” en *Observatorio de contratos públicos*, J. M. Gimeno Feliu, ed., vol. 2018, Thomson Reuters-Aranzadi, 2019, págs. 93-129, isbn: 978-84-1309-943-9.
- [33] J. M. Gimeno Feliu, “Corrupción y seguridad jurídica. La necesidad de un marco normativo de las decisiones públicas anclado en los principios de Integridad y de Transparencia,” *Revista internacional de transparencia e integridad*, n.º 9, pág. 10, 2019. dirección: <https://dialnet.unirioja.es/servlet/articulo?codigo=6977097>.
- [34] A. Zuiderwijk, N. Helbig, J. R. Gil-García y M. Janssen, “Special issue on innovation through open data - A review of the state-of-the-art and an emerging research agenda: Guest editors’ introduction,” *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 9, n.º 2, 2014, issn: 07181876. doi: [10.4067/S0718-18762014000200001](https://doi.org/10.4067/S0718-18762014000200001).
- [35] D. Corrales-Garay, M. Ortiz-de-Urbina-Criado y E. M. Mora-Valentín, “Knowledge areas, themes and future research on open data: A co-word analysis,” *Government Information Quarterly*, vol. 36, n.º 1, págs. 77-87, 2018, issn: 0740624X. doi: [10.1016/j.giq.2018.10.008](https://doi.org/10.1016/j.giq.2018.10.008).
- [36] J. D. Twizeyimana y A. Andersson, “The public value of E-Government – A literature review,” *Government Information Quarterly*, vol. 36, n.º 2, págs. 167-178, 2019, issn: 0740-624X. doi: [10.1016/J.GIQ.2019.01.001](https://doi.org/10.1016/J.GIQ.2019.01.001).
- [37] G. Magalhaes y C. Roseira, “Open government data and the private sector: An empirical view on business models and value creation,” *Government Information*

- Quarterly, n.º August, págs. 1-10, 2017, issn: 0740624X. doi: [10.1016/j.giq.2017.08.004](https://doi.org/10.1016/j.giq.2017.08.004).
- [38] J. M. Alvarez-Rodríguez, J. E. Labra-Gayo y P. O. De Pablos, “New trends on e-Procurement applying semantic technologies: Current status and future challenges,” *Computers in Industry*, vol. 65, n.º 5, págs. 800-820, 2014, issn: 01663615. doi: [10.1016/j.compind.2014.04.005](https://doi.org/10.1016/j.compind.2014.04.005).
- [39] E. Huyer y L. van Knippenberg, “The Economic Impact of Open Data Opportunities for value creation in Europe,” European Commission, inf. téc., 2020, pág. 138. doi: [10.2830/63132](https://doi.org/10.2830/63132).
- [40] W. McKinney, *Python for Data Analysis 2ed*, Second, M. Beaugureau, ed. O’Reilly, 2012, pág. 541, isbn: 978-1-491-95766-0.
- [41] M. Lutz, *Learning Python*, 5.^a ed. O’Reilly, 2013, pág. 1594, isbn: 9781449355739.
- [42] A. Oussous, F.-z. Benjelloun, A. Ait y S. Belfkih, “Big Data technologies: A survey,” *Journal of King Saud University - Computer and Information Sciences*, vol. 30, n.º 4, págs. 431-448, 2018, issn: 1319-1578. doi: [10.1016/j.jksuci.2017.06.001](https://doi.org/10.1016/j.jksuci.2017.06.001).
- [43] J. Miranzo Díaz, “El nuevo Derecho de la UE: las medidas anticorrupción en la contratación pública,” Tesis doct., Universidad de Castilla-La Mancha, 2018, pág. 568.
- [44] E. Díaz Bravo y J. A. Moreno Molina, *Contratación Pública Global: Visiones Comparadas*. Valencia: Tirant lo Blanch, 2020, pág. 782, isbn: 9788413367279.
- [45] A. Huergo Lora, “Administraciones Públicas e inteligencia artificial: ¿más o menos discrecionalidad?” *El Cronista del Estado Social y Democrático de Derecho*, vol. 96-97, págs. 78-95, 2021, issn: 1889-0016.
- [46] J. D. Kelleher, B. Mac Namee y A. D’Arcy, *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*, First edit. The MIT Press, 2015, pág. 624, isbn: 9780262029445.
- [47] I. Lee e Y. J. Shin, “Machine learning for enterprises: Applications, algorithm selection, and challenges,” *Business Horizons*, vol. 63, n.º 2, págs. 157-170, 2020, issn: 00076813. doi: [10.1016/j.bushor.2019.10.005](https://doi.org/10.1016/j.bushor.2019.10.005).
- [48] World Bank, “Disruptive technologies in public procurement,” The World Bank, Washington, inf. téc., 2021, pág. 114. dirección: <https://documents1.worldbank.org/curated/en/522181612428427520/pdf/Disruptive-Technologies-in-Public-Procurement.pdf>.
- [49] S. Athey y G. Imbens, “Machine Learning Methods Economists Should Know About,” *Working Paper*, n.º March, mar. de 2019. arXiv: 1903.10075. dirección: <https://www.gsb.stanford.edu/faculty-research/working-papers/machine-learning-methods-economists-should-know-about%20http://arxiv.org/abs/1903.10075>.

- [50] S. Mullainathan y J. Spiess, “Machine Learning: An Applied Econometric Approach,” *Journal of Economic Perspectives*, vol. 31, n.º 2, págs. 87-106, 2017, issn: 0895-3309. doi: [10.1257/jep.31.2.87](https://doi.org/10.1257/jep.31.2.87).
- [51] H. R. Varian, “Big Data: New Tricks for Econometrics,” *Journal of Economic Perspectives*, vol. 28, n.º 2, págs. 3-28, 2014, issn: 0895-3309. doi: [10.1257/jep.28.2.3](https://doi.org/10.1257/jep.28.2.3).
- [52] T. Hastie, R. Tibshirani y J. Friedman, *The Elements of Statistical Learning* (Springer Series in Statistics), Second. New York, NY: Springer New York, 2009, pág. 764, isbn: 978-0-387-84857-0. doi: [10.1007/978-0-387-84857-0](https://doi.org/10.1007/978-0-387-84857-0).
- [53] K. P. Murphy, *Probabilistic ML - Adaptive Computation and Machine Learning*. 2022, isbn: 2021027430. dirección: <https://lccn.loc.gov/2021027430>.
- [54] A. Soylu y col., “Data Quality Barriers for Transparency in Public Procurement,” *Information*, vol. 13, n.º 2, pág. 99, feb. de 2022, issn: 2078-2489. doi: [10.3390/info13020099](https://doi.org/10.3390/info13020099).
- [55] M. M. Khurshid, N. H. Zakaria, A. Rashid, M. N. Ahmad, M. I. Arfeen y H. M. Faisal Shehzad, “Modeling of Open Government Data for Public Sector Organizations Using the Potential Theories and Determinants—A Systematic Review,” *Informatics*, vol. 7, n.º 3, pág. 24, jul. de 2020, issn: 2227-9709. doi: [10.3390/informatics7030024](https://doi.org/10.3390/informatics7030024).
- [56] S. de Juana-Espinosa y S. Luján-Mora, “Open government data portals in the European Union: Considerations, development, and expectations,” *Technological Forecasting and Social Change*, vol. 149, n.º October, pág. 119 769, 2019, issn: 00401625. doi: [10.1016/j.techfore.2019.119769](https://doi.org/10.1016/j.techfore.2019.119769).
- [57] G. Vancauwenberghe, B. van Loenen y J. Cromptvoets, *Open Data Exposed*. T.M.C. Asser Press, The Hague, 2018, pág. 299, isbn: 9789462652606. doi: <https://doi.org/10.1007/978-94-6265-261-3>.
- [58] S. Sadiq y M. Indulska, “Open data: Quality over quantity,” *International Journal of Information Management*, vol. 37, n.º 3, págs. 150-154, 2017, issn: 02684012. doi: [10.1016/j.ijinfomgt.2017.01.003](https://doi.org/10.1016/j.ijinfomgt.2017.01.003).
- [59] Federación Española de Municipios y Provincias, ed., *Datos Abiertos: Guía estratégica para su puesta en marcha Conjuntos de datos mínimos a publicar*, 1.ª ed. Spain: Wolters Kluwer España, S.A, 2017, pág. 122, isbn: 9788415651574.
- [60] F. Ahmadi Zeleti, A. Ojo y E. Curry, “Exploring the economic value of open government data,” *Government Information Quarterly*, vol. 33, n.º 3, págs. 535-551, 2016, issn: 0740624X. doi: [10.1016/j.giq.2016.01.008](https://doi.org/10.1016/j.giq.2016.01.008).

- [61] J. Attard, F. Orlandi, S. Scerri y S. Auer, "A systematic review of open government data initiatives," *Government Information Quarterly*, vol. 32, n.º 4, págs. 399-418, 2015, issn: 0740624X. doi: [10.1016/j.giq.2015.07.006](https://doi.org/10.1016/j.giq.2015.07.006).
- [62] D. Goens, "The exploitation of Business Register data from a public sector information and data protection perspective: A case study," *Computer Law & Security Review*, vol. 26, n.º 4, págs. 398-405, jul. de 2010, issn: 02673649. doi: [10.1016/j.clsr.2010.05.001](https://doi.org/10.1016/j.clsr.2010.05.001).
- [63] M. Kuziemski y G. Misuraca, "AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings," *Telecommunications Policy*, vol. 44, n.º 6, pág. 13, jul. de 2020, issn: 03085961. doi: [10.1016/j.telpol.2020.101976](https://doi.org/10.1016/j.telpol.2020.101976).
- [64] N. Obwegeser y S. D. Müller, "Innovation and public procurement: Terminology, concepts, and applications," *Technovation*, vol. 74-75, n.º April 2016, págs. 1-17, 2018, issn: 01664972. doi: [10.1016/j.technovation.2018.02.015](https://doi.org/10.1016/j.technovation.2018.02.015).
- [65] M. C. B. de Araújo, L. H. Alencar y C. M. de Miranda Mota, "Project procurement management: A structured literature review," *International Journal of Project Management*, vol. 35, n.º 3, págs. 353-377, 2017. doi: [10.1016/j.ijproman.2017.01.008](https://doi.org/10.1016/j.ijproman.2017.01.008).
- [66] K. V. Thai, ed., *Global Public Procurement Theories and Practices*. Springer International Publishing, 2017, isbn: 978-3-319-49279-7. doi: [10.1007/978-3-319-49280-3](https://doi.org/10.1007/978-3-319-49280-3).
- [67] A. S. Patrucco, D. Luzzini y S. Ronchi, "Research perspectives on public procurement: Content analysis of 14 years of publications in the journal of public procurement," *Journal of Public Procurement*, vol. 17, n.º 2, págs. 229-269, mar. de 2017, issn: 1535-0118. doi: [10.1108/JOPP-17-02-2017-B003](https://doi.org/10.1108/JOPP-17-02-2017-B003).
- [68] M. K. Gorgun, M. Kutlu y B. K. Onur Tas, "Predicting The Number of Bidders in Public Procurement," en *2020 5th International Conference on Computer Science and Engineering (UBMK)*, IEEE, sep. de 2020, págs. 360-365, isbn: 978-1-7281-7565-2. doi: [10.1109/UBMK50275.2020.9219404](https://doi.org/10.1109/UBMK50275.2020.9219404).
- [69] M. Bilal y L. O. Oyedele, "Big Data with deep learning for benchmarking profitability performance in project tendering," *Expert Systems with Applications*, vol. 147, pág. 113 194, 2020, issn: 09574174. doi: [10.1016/j.eswa.2020.113194](https://doi.org/10.1016/j.eswa.2020.113194).
- [70] J. M. Kim y H. Jung, "Predicting bid prices by using machine learning methods," *Applied Economics*, vol. 51, n.º 19, págs. 2011-2018, 2019, issn: 14664283. doi: [10.1080/00036846.2018.1537477](https://doi.org/10.1080/00036846.2018.1537477).

- [71] R. Matin, C. Hansen, C. Hansen y P. Mølgaard, “Predicting distresses using deep learning of text segments in annual reports,” *Expert Systems With Applications*, vol. 132, págs. 199-208, 2019, issn: 0957-4174. doi: [10.1016/j.eswa.2019.04.071](https://doi.org/10.1016/j.eswa.2019.04.071).
- [72] J. S. Chou, C. W. Lin, A. D. Pham y J. Y. Shao, “Optimized artificial intelligence models for predicting project award price,” *Automation in Construction*, vol. 54, págs. 106-115, 2015, issn: 09265805. doi: [10.1016/j.autcon.2015.02.006](https://doi.org/10.1016/j.autcon.2015.02.006).
- [73] M. Huber y D. Imhof, “Deep learning for detecting bid rigging: Flagging cartel participants based on convolutional neural networks,” abr. de 2021. arXiv: [2104.11142](https://arxiv.org/abs/2104.11142). dirección: <http://arxiv.org/abs/2104.11142>.
- [74] M. E. K. Niessen, J. M. Paciello y J. I. P. Fernandez, “Anomaly Detection in Public Procurements using the Open Contracting Data Standard,” en *2020 Seventh International Conference on eDemocracy & eGovernment (ICEDEG)*, IEEE, abr. de 2020, págs. 127-134, isbn: 978-1-7281-5882-2. doi: [10.1109/ICEDEG48599.2020.9096674](https://doi.org/10.1109/ICEDEG48599.2020.9096674).
- [75] Y. Torres Berru, V. F. López Batista, P. Torres-Carrión y M. G. Jimenez, “Artificial Intelligence Techniques to Detect and Prevent Corruption in Procurement: A Systematic Literature Review,” en *Communications in Computer and Information Science*, Springer, ed., vol. 2, 2020, págs. 254-268, isbn: 9783030425203. doi: [10.1007/978-3-030-42520-3_21](https://doi.org/10.1007/978-3-030-42520-3_21).
- [76] M. Huber, D. Imhof y R. Ishii, “Transnational machine learning with screens for flagging bid-rigging cartels,” *University of Fribourg (Switzerland)*, vol. Working Pa, n.º Faculty of Economics and Social Sciences, 2020. dirección: https://doc.rero.ch/record/329575/files/WP%7B%5C_%7DSES%7B%5C_%7D519.pdf.
- [77] H. Wallimann, D. Imhof y M. Huber, “A Machine Learning Approach for Flagging Incomplete Bid-rigging Cartels,” *University of Freiburg/Fribourg (Switzerland)*, abr. de 2020. arXiv: [2004.05629](https://arxiv.org/abs/2004.05629). dirección: <http://arxiv.org/abs/2004.05629>.
- [78] M. Lima, R. Silva, F. Lopes de Souza Mendes, L. R. de Carvalho, A. Araujo y F. de Barros Vidal, “Inferring about fraudulent collusion risk on Brazilian public works contracts in official texts using a Bi-LSTM approach,” en *Findings of the Association for Computational Linguistics: EMNLP 2020*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, págs. 1580-1588. doi: [10.18653/v1/2020.findings-emnlp.143](https://doi.org/10.18653/v1/2020.findings-emnlp.143).
- [79] C. Giorgiantonio y F. Decarolis, “Corruption Red Flags in Public Procurement: New Evidence from Italian Calls for Tenders,” *SSRN Electronic Journal*, n.º February, pág. 34, 2020, issn: 1556-5068. doi: [10.2139/ssrn.3612661](https://doi.org/10.2139/ssrn.3612661).

- [80] J. Gallego, G. Rivero y J. Martínez, “Preventing rather than punishing: An early warning model of malfeasance in public procurement,” *International Journal of Forecasting*, jul. de 2020, issn: 01692070. doi: [10.1016/j.ijforecast.2020.06.006](https://doi.org/10.1016/j.ijforecast.2020.06.006).
- [81] M. Huber y D. Imhof, “Machine learning with screens for detecting bid-rigging cartels,” *International Journal of Industrial Organization*, vol. 65, págs. 277-301, jul. de 2019, issn: 01677187. doi: [10.1016/j.ijindorg.2019.04.002](https://doi.org/10.1016/j.ijindorg.2019.04.002).
- [82] K. Rabuzin y N. Modrušan, “Prediction of Public Procurement Corruption Indices using Machine Learning Methods,” en *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, SCITEPRESS - Science y Technology Publications, 2019, págs. 333-340, isbn: 978-989-758-382-7. doi: [10.5220/0008353603330340](https://doi.org/10.5220/0008353603330340).
- [83] J. Wacker, R. P. Ferreira y M. Ladeira, “Detecting Fake Suppliers using Deep Image Features,” en *2018 7th Brazilian Conference on Intelligent Systems (BRACIS)*, IEEE, oct. de 2018, págs. 224-229, isbn: 978-1-5386-8023-0. doi: [10.1109/BRACIS.2018.00046](https://doi.org/10.1109/BRACIS.2018.00046).
- [84] T. Sun y L. J. Sales, “Predicting Public Procurement Irregularity: An Application of Neural Networks,” *Journal of Emerging Technologies in Accounting*, vol. 15, n.º 1, págs. 141-154, jul. de 2018, issn: 1558-7940. doi: [10.2308/jeta-52086](https://doi.org/10.2308/jeta-52086).
- [85] M. Lei, Z. Yin, S. Li y H. Li, “Detecting the collusive bidding behavior in below average bid auction,” en *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, IEEE, jul. de 2017, págs. 1720-1727, isbn: 978-1-5386-2165-3. doi: [10.1109/FSKD.2017.8393026](https://doi.org/10.1109/FSKD.2017.8393026).
- [86] E. Grace, A. Rai, E. Redmiles y R. Ghani, “Detecting fraud, corruption, and collusion in international development contracts: The design of a proof-of-concept automated system,” en *2016 IEEE International Conference on Big Data (Big Data)*, IEEE, dic. de 2016, págs. 1444-1453, isbn: 978-1-4673-9005-7. doi: [10.1109/BigData.2016.7840752](https://doi.org/10.1109/BigData.2016.7840752).
- [87] D. Imhof, “Empirical Methods for Detecting Bid-rigging Cartels,” Tesis doct., Université Bourgogne Franche-Comté, 2018. dirección: <https://tel.archives-ouvertes.fr/tel-01963076>.
- [88] F. Pedregosa y col., “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, págs. 2825-2830, 2011. arXiv: [1201.0490](https://arxiv.org/abs/1201.0490). dirección: <http://arxiv.org/abs/1201.0490>.
- [89] A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 1.ª ed. O'Reilly, 2017, isbn: 978-1-491-96229-9.

- [90] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann e I. H. Witten, "The WEKA data mining software," *ACM SIGKDD Explorations Newsletter*, vol. 11, n.º 1, págs. 10-18, nov. de 2009, issn: 1931-0145. doi: [10.1145/1656274.1656278](https://doi.org/10.1145/1656274.1656278).
- [91] I. H. Witten, E. Frank y M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, Third. Boston: Elsevier, 2011, isbn: 9780123748560. doi: [10.1016/C2009-0-19715-5](https://doi.org/10.1016/C2009-0-19715-5).
- [92] M. Abadi y col., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," mar. de 2016. arXiv: [1603.04467](https://arxiv.org/abs/1603.04467). dirección: <http://arxiv.org/abs/1603.04467>.
- [93] L. Breiman, "Random forests," *Machine Learning*, vol. 45, n.º 1, págs. 5-32, 2001. doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- [94] A. Verikas, A. Gelzinis y M. Bacauskiene, "Mining data with random forests: A survey and results of new tests," *Pattern Recognition*, vol. 44, n.º 2, págs. 330-349, 2011, issn: 00313203. doi: [10.1016/j.patcog.2010.08.011](https://doi.org/10.1016/j.patcog.2010.08.011).
- [95] M. Fernández-Delgado, M. S. Sirsat, E. Cernadas, S. Alawadi, S. Barro y M. Febrero-Bande, "An extensive experimental survey of regression methods," *Neural Networks*, vol. 111, págs. 11-34, 2019, issn: 18792782. doi: [10.1016/j.neunet.2018.12.010](https://doi.org/10.1016/j.neunet.2018.12.010).
- [96] M. R. Segal, "Machine Learning Benchmarks and Random Forest Regression," *UCSF: Center for Bioinformatics and Molecular Biostatistics*, 2004. dirección: <https://escholarship.org/uc/item/35x3v9t4>.
- [97] G. Biau y E. Scornet, "A random forest guided tour," *Test*, vol. 25, n.º 2, págs. 197-227, 2016, issn: 11330686. doi: [10.1007/s11749-016-0481-7](https://doi.org/10.1007/s11749-016-0481-7).
- [98] A. Mehrbod y A. Grilo, "Tender calls search using a procurement product named entity recogniser," *Advanced Engineering Informatics*, vol. 36, n.º June 2017, págs. 216-228, abr. de 2018, issn: 14740346. doi: [10.1016/j.aei.2018.04.005](https://doi.org/10.1016/j.aei.2018.04.005).
- [99] M. Nečáský, J. Klímek, J. Mynarz, T. Knap, V. Svátek y J. Stárka, "Linked data support for filing public contracts," *Computers in Industry*, vol. 65, n.º 5, págs. 862-877, jun. de 2014, issn: 01663615. doi: [10.1016/j.compind.2013.12.006](https://doi.org/10.1016/j.compind.2013.12.006).
- [100] *Algorithms and Collusion: Competition Policy in the Digital Age*, 2017. dirección: www.oecd.org/competition/algorithms-collusion-competition-policy-in-the-digital-age.htm.
- [101] D. Imhof, Y. Karagök y S. Rutz, "Screening for Bid Rigging - Does it works?" *Journal of Competition Law & Economics*, vol. 14, n.º 2, págs. 235-261, jun. de 2018, issn: 1744-6414. doi: [10.1093/joclec/nhy006](https://doi.org/10.1093/joclec/nhy006).

- [102] D. Imhof, "Simple Statistical Screens to Detect Bid Rigging," *Working Papers SES. Faculty of Economics and Social Sciences. University of Fribourg*, vol. 484, 2017. dirección: http://doc.rero.ch/record/289133/files/WP_SES_484.pdf.
- [103] T. Zhang, "Solving large scale linear prediction problems using stochastic gradient descent algorithms," en *Twenty-first international conference on Machine learning - ICML '04*, New York, New York, USA: ACM Press, 2004, pág. 116, isbn: 1581138285. doi: [10.1145/1015330.1015332](https://doi.org/10.1145/1015330.1015332).
- [104] P. Geurts, D. Ernst y L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, n.º 1, págs. 3-42, 2006, issn: 08856125. doi: [10.1007/s10994-006-6226-1](https://doi.org/10.1007/s10994-006-6226-1).
- [105] Y. Freund y R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences*, vol. 55, n.º 1, págs. 119-139, ago. de 1997, issn: 00220000. doi: [10.1006/jcss.1997.1504](https://doi.org/10.1006/jcss.1997.1504).
- [106] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," *The Annals of Statistics*, vol. 29, n.º 5, págs. 1189-1232, oct. de 2001, issn: 0090-5364. doi: [10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451).
- [107] C. Cortes y V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, n.º 3, págs. 273-297, sep. de 1995, issn: 0885-6125. doi: [10.1007/BF00994018](https://doi.org/10.1007/BF00994018).
- [108] N. S. Altman, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression," *The American Statistician*, vol. 46, n.º 3, págs. 175-185, ago. de 1992, issn: 0003-1305. doi: [10.1080/00031305.1992.10475879](https://doi.org/10.1080/00031305.1992.10475879).
- [109] C. E. Rasmussen y C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006, isbn: 9780262182539. dirección: <http://gaussianprocess.org/gpml>.
- [110] S. Curto, S. Ghislandi, K. Van de Vooren, S. Duranti y L. Garattini, "Regional tenders on biosimilars in Italy: An empirical analysis of awarded prices," *Health Policy*, vol. 116, n.º 2-3, págs. 182-187, 2014, issn: 18726054. doi: [10.1016/j.healthpol.2014.02.011](https://doi.org/10.1016/j.healthpol.2014.02.011).
- [111] T. Hanák y P. Muchová, "Impact of Competition on Prices in Public Sector Procurement," *Procedia Computer Science*, vol. 64, págs. 729-735, 2015, issn: 18770509. doi: [10.1016/j.procs.2015.08.601](https://doi.org/10.1016/j.procs.2015.08.601).
- [112] J. Soudek y J. Skuhrovec, "Procurement procedure, competition and final unit price: The case of commodities," *Journal of Public Procurement*, vol. 16, n.º 1, págs. 1-21, mar. de 2016, issn: 1535-0118. doi: [10.1108/JOPP-16-01-2016-B001](https://doi.org/10.1108/JOPP-16-01-2016-B001).

Enlaces a internet

- Autoridad Independiente de Responsabilidad Fiscal (AIReF): <https://www.airef.es>
- Autoridad Suiza de la Competencia: <https://www.weko.admin.ch/weko/en/home/comco.html>
- Base Nacional de Subvenciones: <https://www.pap.hacienda.gob.es/bdnstrans/es/index>
- Central de Información Económico-Financiera de las AA.PP.: <https://www.hacienda.gob.es/es-ES/CDI/Paginas/centraldeinformacion.aspx>
- Comisión Nacional de los Mercados y la Competencia (CNMC): <http://www.cnmc.es>
- Consejo de Transparencia y Buen Gobierno (CTBG): <https://www.consejodetransparencia.es>
- Diario Oficial de la Unión Europea (DOUE): <https://eur-lex.europa.eu/oj/direct-access.html>
- Dirección General de Racionalización y Centralización de la Contratación (DGRCC): <https://contratacioncentralizada.gob.es>
- DoZorro: <https://dozorro.org>
- Guardia Civil: <https://www.guardiacivil.es>
- Intervención General de la Administración del Estado (IGAE): <https://www.igae.pap.hacienda.gob.es/sitios/igae>
- Junta Consultiva de Contratación del Estado: <https://www.hacienda.gob.es/es-ES/Areas%20Tematicas/Contratacion/Junta%20Consultiva%20de%20Contratacion%20Administrativa/Paginas/default.aspx>
- Oficina Independiente de Regulación y Supervisión de la Contratación (OIReScon): <https://www.hacienda.gob.es/es-ES/Oirescon/Paginas/HomeOirescon.aspx>
- Open Knowledge Foundation: <https://okfn.org>
- Open Tender: <https://www.opentender.eu>
- Plataforma de Contratación del Sector Público (PLACSP): <https://contrataciondelestado.es>
- Policía Nacional: <https://www.policia.es>
- Policía Federal de Brasil: <https://www.gov.br/pf/pt-br>
- Portal de datos abiertos de la UE: <https://data.europa.eu/>
- Portal de la Transparencia de la AGE: <https://transparencia.gob.es>

- Red flags: <https://www.redflags.eu>
- Registro de Contratos: <https://www.hacienda.gob.es/es-ES/Areas%20Tematicas/Contratacion/Junta%20Consultiva%20de%20Contratacion%20Administrativa/Paginas/Registro%20publico%20de%20contratos.aspx>
- Registro Mercantil: <https://www.registradores.org>
- Tenders Electronic Daily (TED): <https://ted.europa.eu/>
- Transparencia Internacional: <https://www.transparency.org>
- Tribunal Administrativo Central de Recursos Contractuales: <https://www.hacienda.gob.es/es-ES/Areas%20Tematicas/Contratacion/TACRC/Paginas/Tribunal%20Administrativo%20Central%20de%20Recursos%20Contractuales.aspx>
- Tribunal de Cuentas: <https://www.tcu.es>

ISBN: 978-84-8333-724-0



9 788483 337240