AÑO 4. No. 3
ISSN:2992-7978

DETECCIÓN DE INDICIOS DE RIESGO EN CONTRATACIONES MUNICIPALES MEDIANTE ANALÍTICA DE DATOS: UN ENFOQUE EXPLORATORIO

Miguel Angel Chávez Herrera[1]
ORCID: 0000-0002-9391-6231

Rosa Gabriela Camero Berrones[2]
ORCID: 0000-0003-4438-1645

[1] Doctorando, Universidad Americana de Europa (UNADE). Correo: migchavezh@gmail.com
[2] Profesora-investigadora, Universidad Americana de Europa (UNADE)

RESUMEN

La investigación se propuso identificar indicios analíticos de riesgo en los procesos de contratación pública de 20 gobiernos locales peruanos mediante técnicas de analítica de datos. El estudio adoptó un enfoque cuantitativo, de diseño no experimental y transversal, con un alcance descriptivo y exploratorio. Se empleó el modelo CRISP-DM como marco metodológico y se implementó un proceso de Extracción, Transformación y Carga (ETL) mediante scripts SQL en PostgreSQL para integrar datos del Portal de Datos Abiertos del OSCE del periodo 2022–2024. Los resultados permitieron identificar patrones de riesgo asociados a la concentración financiera de proveedores (hasta 45.48% en un solo actor), indicios de fraccionamiento contractual que superaron el umbral de 8 UIT y recurrencias atípicas entre comités de selección y proveedores (hasta 18 coincidencias). Se concluyó que el modelo analítico propuesto es una herramienta eficaz para la detección de áreas de riesgo, no obstante, es imperativo diferenciar la 'correlación analítica' de la 'inferencia de corrupción'. Los hallazgos presentados constituyen indicios objetivos de riesgo (red flags) que señalan anomalías administrativas, pero su calificación como actos de corrupción requiere de una validación externa y un debido proceso legal por parte de los órganos de control.

Palabras clave: Analítica de Datos, Contratación Pública, Gestión de Riesgos.

ABSTRACT

The research aimed to identify analytical risk indicators in the public procurement processes of 20 Peruvian local governments through data analytics techniques. The study adopted a quantitative, non-experimental, cross-sectional design with a descriptive and exploratory scope. The CRISP-DM model was used as a methodological framework, and an Extraction, Transformation, and Loading (ETL) process was implemented using SQL scripts in PostgreSQL to integrate data from the OSCE Open Data Portal for the 2022–2024 period. The results identified risk patterns associated with financial concentration of suppliers (up to 45.48% in a single actor), indications of contractual splitting exceeding the 8 UIT threshold, and atypical recurrences between selection committees and suppliers (up to 18 coincidences). It was concluded that the proposed analytical model is an effective tool to detect risk areas; however, it is imperative to differentiate 'analytical correlation' from 'inference of corruption'. The findings constitute objective risk indicators (red flags) that point to administrative anomalies, but their qualification as corrupt acts requires external validation and due legal process by control bodies.

Keywords: Data Analytics, Public Procurement, Risk Management.

AÑO 4. No. 3
ISSN:2992-7978

INTRODUCCIÓN

La contratación pública en el Perú representa un área crítica para la transparencia debido al alto volumen presupuestal y la multiplicidad de actores involucrados. El Organismo Supervisor de las Contrataciones del Estado (OSCE) publica datos masivos, pero su potencial analítico es subutilizado debido a la fragmentación de la información. El objetivo general es identificar indicios analíticos de riesgo en municipalidades mediante analítica descriptiva, por lo que el estudio busca responder: ¿en qué medida la aplicación de analítica descriptiva permite identificar patrones de riesgo en las compras municipales? Esta investigación se justifica por la necesidad de migrar de un análisis estático a un modelo dinámico de detección de anomalías.

ESTADO DEL ARTE

La literatura destaca que los datos históricos de compras públicas contienen patrones que revelan riesgos de corrupción e ineficiencias, es así que García-Rodríguez et al. (2025) proponen modelos de aprendizaje automático para clasificar riesgos, mientras que RedeIA (2021) enfatiza el uso de indicadores de concentración y fraccionamiento. En el Perú, esfuerzos previos como los de Proética (2021) han señalado barreras técnicas en el acceso y calidad de los datos del OSCE, limitando la supervisión oportuna, hallazgos que coinciden con lo planteado en el ámbito latinoamericano por Jorquera (2019), Felizzola et al. (2024) y Muñoz-Cancino y Ríos (2025), quienes demuestran que el procesamiento masivo de datos relacionales y el uso de técnicas analíticas avanzadas en contrataciones estatales permiten identificar patrones anómalos de concentración y direccionamiento que las auditorías gubernamentales tradicionales suelen pasar por alto.

MARCO TEÓRICO

La investigación se sustenta en tres pilares teóricos y una metodología estándar para el tratamiento de datos. Al respecto, en primer lugar, la Teoría de la Agencia proporciona el marco conceptual para comprender los problemas de asimetría de información y conflictos de interés en las contrataciones públicas, donde la ciudadanía (principal) delega en los funcionarios públicos (agentes) la responsabilidad de gestionar recursos estatales, donde los riesgos identificados surgen cuando los agentes tienen incentivos para tomar decisiones que beneficien intereses personales en detrimento del interés público, especialmente cuando los mecanismos de supervisión son limitados, contando con la transparencia y los sistemas de control como mecanismos que actúan para alinear los incentivos de los agentes con los objetivos del principal, reduciendo los costos de agencia.

En segundo lugar, la Economía de la Corrupción analiza las condiciones que favorecen prácticas irregulares, siendo que los modelos teóricos sugieren que estas conductas florecen ante la alta discrecionalidad de los funcionarios, la existencia de rentas significativas capturables y una baja probabilidad de detección y sanción, lo que en las compras municipales, esto se manifiesta en el diseño discrecional de especificaciones técnicas y la complejidad de procesos que dificultan el monitoreo exhaustivo, por ello, el modelo analítico propuesto busca incrementar la probabilidad de detección mediante el monitoreo sistemático basado en datos históricos, elevando así el costo esperado de incurrir en prácticas corruptas.

En tercer lugar, la Teoría de la Decisión y el Análisis de Riesgos permiten abordar la incertidumbre en los procesos contractuales ya que el análisis de riesgos, mediante técnicas cuantitativas, estima la probabilidad de ocurrencia de eventos adversos, facilitando la priorización de acciones preventivas. Los sistemas de detección operacionalizan estos principios al calcular indicadores de riesgo, establecer umbrales de alerta y generar señales para la intervención de los órganos de supervisión.

AÑO 4. No. 3
ISSN:2992-7978

Por el lado metodológico, para el estudio se aplica el modelo CRISP-DM (Cross Industry Standard Process for Data Mining), el cual es un marco que estructura el ciclo analítico de minería de datos en seis fases interrelacionadas: Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación y Despliegue, donde el carácter iterativo de esta metodología permite retornar a fases anteriores conforme se adquiere nuevo conocimiento durante el desarrollo del proyecto, asegurando un análisis riguroso y reproducible, siendo pertinente precisar que, en este marco metodológico, se debe tener en cuenta que la analítica descriptiva agrega y cuantifica los datos históricos, la minería de datos extrae patrones no evidentes mediante consultas estructuradas en SQL (como vistas lógicas parametrizadas), y la detección de anomalías operacionaliza las alertas rojas (red flags) cuando los indicadores superan los umbrales normativos y estadísticos definidos.

Finalmente, para efectos de esta investigación y en alineación con el marco normativo e instrumental de las contrataciones del Estado, se establecen las siguientes definiciones operacionales y conceptuales:

Indicio de Riesgo: Representa una señal atípica o patrón anómalo cuantitativo en los registros administrativos que, sin constituir una prueba fehaciente de ilegalidad por sí misma, señala una desviación respecto a la conducta de mercado esperada y sugiere una mayor probabilidad de vulneración a los principios de transparencia y eficiencia.

Riesgo Relacional: Fenómeno analizado desde la Teoría de la Agencia, definido como la presencia reiterada y estadísticamente atípica de coincidencias entre los mismos evaluadores (miembros del comité de selección) y los mismos proveedores adjudicados en múltiples procesos. Este indicador advierte un probable conflicto de interés, captura del regulador o sesgo de parcialidad en la toma de decisiones.

Fraccionamiento: Es la práctica prohibida consistente en dividir de forma artificial la contratación de bienes, servicios u obras con la finalidad de eludir el procedimiento de selección competitivo que corresponde según los umbrales legales (Art. 13 del T.U.O. de la Ley N.º 30225), la misma que en esta investigación se operacionaliza mediante consultas SQL en PostgreSQL sobre la tabla de contratos, agrupando por entidad y proveedor para identificar la emisión recurrente de órdenes dentro de un mismo año fiscal. El algoritmo parametriza el valor legal de la UIT para cada periodo fiscal (S/ 4,600 en 2022; S/ 4,950 en 2023; y S/ 5,150 en 2024), detectando la acumulación de compras directas cuya suma agregada supera las 8 UIT y cuyo valor promedio por orden bordea dicho límite, evidenciando riesgo de elusión de procesos competitivos.

Pluralidad de Postores: Es la garantía y el principio rector que busca asegurar la máxima participación de proveedores competidores calificados en los procesos de contratación pública con la finalidad de evitar barreras burocráticas o direccionamientos que restrinjan la libre competencia, permitiendo al Estado obtener la propuesta más ventajosa mediante la comparación de múltiples ofertas válidas (Art. 2 del T.U.O. de la Ley N.º 30225). Se operacionaliza mediante consultas SQL en PostgreSQL cruzando las tablas de procedimientos de selección, participantes registrados y ofertas presentadas / admitidas por cada ítem de un proceso en un año fiscal específico. El algoritmo evalúa el nivel de competencia efectiva mediante un ratio de concurrencia y la concentración de adjudicaciones, identificando procesos críticos con "postor único" o con un número de ofertas inferior al promedio sectorial.

METODOLOGÍA

El estudio adoptó un enfoque cuantitativo y un diseño no experimental de corte transversal, con un alcance descriptivo y exploratorio en la que se utilizó la analítica de datos para examinar el comportamiento

AÑO 4. No. 3
ISSN:2992-7978

histórico de las compras públicas a partir de registros administrativos existentes, sin manipular las variables de estudio.

La población estuvo constituida por la totalidad de registros de contratación pública de las municipalidades peruanas disponibles en el Portal de Datos Abiertos del Organismo Supervisor de las Contrataciones del Estado (OSCE) durante el periodo 2022–2024, de la que se seleccionó una muestra de veinte (20) entidades municipales de los departamentos de Lima y Áncash mediante un muestreo no probabilístico intencional basado en tres criterios técnicos: (i) representatividad de zonas de alta actividad contractual (departamentos de Lima y Áncash), (ii) integridad de los registros en el Portal de Datos Abiertos del OSCE durante el periodo 2022–2024, asegurando la disponibilidad de tablas de convocatorias, adjudicaciones, contratos, proveedores y comités; y (iii) capacidad de validación cruzada, seleccionando entidades con trazabilidad completa en el Buscador Público del SEACE. Cabe precisar que a pesar que el tamaño de la muestra no buscó representatividad estadística nacional, los procedimientos y scripts desarrollados fueron diseñados para ser reproducibles y escalables a la totalidad de los gobiernos locales del país.

Respecto de las técnicas e instrumentos de recolección y análisis, se implementó una arquitectura basada en procesos de Extracción, Transformación y Carga (ETL) en la que el instrumento principal fue una matriz de análisis de datos estructurada en una base de datos relacional PostgreSQL, alimentada con archivos en formato CSV y Excel descargados del OSCE. Para el procesamiento de la información, se desarrollaron scripts SQL especializados en la limpieza, normalización y estandarización de datos que originalmente se encontraban fragmentados.

El ciclo analítico se ejecutó siguiendo las seis fases de la metodología CRISP-DM:

Comprensión del negocio: Durante esta fase se definieron los ejes de análisis: concentración de adjudicaciones, patrones de fraccionamiento, pluralidad de postores y evaluación de miembros de comités.

Comprensión de los datos: Se exploraron los datasets del OSCE, identificando tablas de convocatorias, adjudicaciones, contratos, proveedores y comités, evaluando su estructura y calidad inicial.

Preparación de los datos: El pipeline ETL desarrollado superó la fragmentación del OSCE mediante la creación de una arquitectura de datos relacional, en la que se implementaron scripts SQL de limpieza para normalizar la información proveniente de los datasets obtenidos del Portal de Datos Abiertos del OSCE y transformar campos mediante el uso de distintas funciones como es el caso de la función REPLACE que principalmente se utilizó para cambiar campos de tipo texto a numérico o a fecha, entre otros. Cabe señalar que el modelado incluyó la creación de vistas lógicas parametrizadas que integraron las distintas tablas generadas, permitiendo el cálculo de variables compuestas para identificar patrones relacionales que la data cruda del portal no muestra de forma directa. A continuación, se muestra parte del proceso de carga de una de las tablas principales de modelo como es el caso de la tabla contratos, para lo cual inicialmente la información, proveniente del portal de datos abiertos del OSCE, se carga en tablas temporales a partir de las cuales se obtiene la información definitiva que irá en las tablas que se usarán para el proceso de análisis:

AÑO 4. No. 3
ISSN:2992-7978

Figura 1

Proceso de Transferencia de la tabla temporal contratos hacia la tabla definitiva

```sql
INSERT INTO contratos (cod_convo, n_item, monto_total, url)
SELECT
    t.cod_convo,
    t.n_item,
    REPLACE(t.monto_total, ',', '.')::NUMERIC(15,2) AS monto_total,
    t.url_contrato
FROM temp_contratos t
WHERE
    t.cod_convo IS NOT NULL
    AND t.n_item IS NOT NULL
    AND EXISTS ( -- Esta es la clave para filtrar
        SELECT 1
        FROM convocatorias c
        WHERE c.cod_convo = t.cod_convo AND c.n_item = t.n_item
    );
```

Modelado: Se desarrollaron modelos analíticos descriptivos mediante consultas SQL avanzadas para calcular la frecuencia y montos de adjudicación, estableciendo rankings (Top 5 y 10), las coincidencias entre miembros de comité y postores que resultaban favorecidos con las adjudicaciones, indicios de fraccionamiento, así como la cantidad de postores que se presentan a un proceso de selección. A continuación, se presenta un extracto de la estructura lógica y el script en PostgreSQL implementados para la extracción de indicios de fraccionamiento (Figura 2):

Figura 2

Script SQL en PostgreSQL para la detección de indicios de fraccionamiento

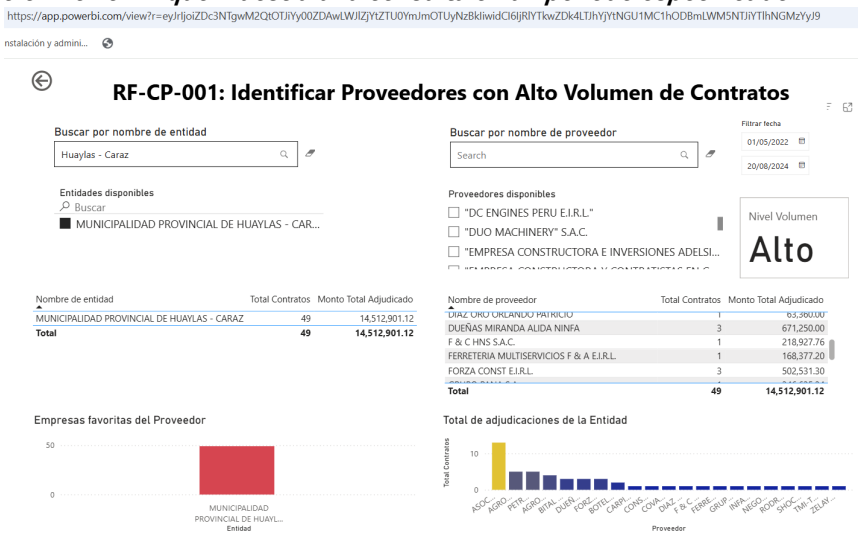
```sql
WITH contratos_base AS (
    SELECT
        EXTRACT(YEAR FROM c.fecha_de_emision) AS anio,
        e.entidad,
        p.nombre_razon_contratista AS proveedor,
        c.nro_de_orden,
        c.monto_total_orden_original AS monto,
        CASE
            WHEN EXTRACT(YEAR FROM c.fecha_de_emision) = 2022 THEN 4600 * 8
            WHEN EXTRACT(YEAR FROM c.fecha_de_emision) = 2023 THEN 4950 * 8
            WHEN EXTRACT(YEAR FROM c.fecha_de_emision) = 2024 THEN 5150 * 8
        END AS limite_8uit
    FROM contrato c
    JOIN entidad e ON c.ruc_entidad = e.ruc_entidad
    JOIN proveedor p ON c.ruc_contratista = p.ruc_contratista
```

Evaluación: La validación del modelo analítico se realizó mediante un proceso de triangulación de datos: en primer lugar, se ejecutó una validación de integridad técnica en PostgreSQL para detectar nulos y duplicados; en segundo lugar, se seleccionaron aleatoriamente hasta cinco (5) registros por municipalidad entre los riesgos identificados de concentración y fraccionamiento, contrastándolos manualmente contra los expedientes digitales en el Buscador Público del SEACE, donde se verificó una correlación del 100% entre los patrones detectados por los scripts SQL y los registros oficiales de adjudicación y contratos, según consta en las respectivas actas de otorgamiento de la buena pro y contratos registrados por el comité de selección en el SEACE, así como las órdenes de compra y servicio registradas por las municipalidades.

Despliegue: Luego que se elaboraron los scripts correspondientes para cada indicador, se generaron vistas para presentar los resultados a través de un prototipo desarrollado en Power BI, en el que se estructuraron dashboards temáticos por tipo de riesgo identificado.

Figura 3

Prototipo en Power BI que muestra una consulta en un periodo especificado



Asimismo, operacionalmente se establecieron cuatro indicadores de riesgo basados en los criterios conceptuales y normativos detallados en el Marco Teórico, cuya síntesis y relación con el marco legal se consolida en la Tabla 1.

Tabla 1

Matriz de variables

Categoría de Análisis (Riesgo/Transparencia)	Definición Operacional (Variables/Indicadores Clave Auditados)	Relación Normativa
Concentración de Proveedores	Porcentaje del monto total adjudicado por el Top 10 de proveedores. Frecuencia de adjudicaciones por proveedor.	Riesgo de vulneración del principio de libre concurrencia y competencia.
Fraccionamiento de Compras	Frecuencia de procesos similares en periodos cortos. Casos donde la suma de órdenes al mismo proveedor supera las 8 UIT en el mismo periodo, siendo cada contratación individual igual o mayor al 80% de dicho umbral.	Práctica prohibida por la Ley de Contrataciones del Estado.
Riesgo de Favorecimiento/Parcialidad	Número de coincidencias repetidas entre miembros de comité y proveedores adjudicados, generando alerta cuando la frecuencia supera el percentil 90 de la distribución observada en la muestra.	Riesgo de conflicto de interés o direccionamiento.
Pluralidad de Postores	Número de postores por proceso de selección.	La Ley de contrataciones del Estado promueve los principios de libre concurrencia, transparencia y pluralidad de postores.

AÑO 4. No. 3
ISSN:2992-7978

RESULTADOS

El análisis de minería de datos sobre el universo de 20 municipalidades seleccionadas (departamentos de Lima y Áncash) durante el periodo 2022–2024 permitió identificar patrones transversales de riesgo en la contratación pública local. En la Tabla 2 se presenta la caracterización cuantitativa agregada del universo analizado, consolidando las métricas descriptivas comparativas por tipo de indicio de riesgo:

Tabla 2

Métricas descriptivas comparativas agregadas del universo de 20 municipalidades analizadas (2022–2024)

Indicador analítico	Cobertura en la Muestra (N=20)	Métrica Agregada / Promedio Muestral	Rango de Variación Observado (Mín. – Máx.)	Caso Emblemático Detectado
Concentración Financiera (% adjudicado al Proveedor Top 1)	60% (12/20)	21.85% del presupuesto ejecutado	8.12% – 45.48%	Huaylas–Caraz: 45.48% (EFF Ingenieros, S/ 42.9 M)
Concentración en la Franja Cercana al Umbral (Monto mayor o igual al 80% de 8 UITs)	85.0% (17/20)	14.3 contrataciones directas por entidad	2 – 92 contrataciones directas por entidad	Casma: 11 contrataciones directas (S/ 388,368.62)
Baja Competencia (% de procesos con 1 solo postor)	90.0% (18/20)	58.4% de procesos sin competencia	22.1% – 85.5%	San Marcos: 85.5% (676 procesos con 1 postor)
Recurrencia Relacional (Coincidencias Comité–Proveedor)	65.0% (13/20)	4.2 coincidencias atípicas / entidad	3 – 18 coincidencias repetidas	San Marcos: 18 coincidencias (E. Romani – A. Antúnez)

Como se muestra en la Tabla 2, se identifican distintos niveles de riesgo asociados a los procesos de contratación pública en las municipalidades analizadas. Cabe precisar que para el indicador analítico de Concentración Financiera, se consideró a las municipalidades cuyo proveedor Top 1 concentra un monto igual o mayor al 20% del monto adjudicado.

Concentración de proveedores: El análisis de la concentración financiera permitió identificar a las entidades con mayor dependencia hacia actores específicos. Bajo el criterio técnico de una participación individual superior al 30% del monto total adjudicado, se identificaron cinco municipalidades en situación de riesgo crítico, tal como se detalla en la Tabla 3.

Tabla 3

Nivel de concentración de proveedores por municipalidad (Top 5, 2022–2024)

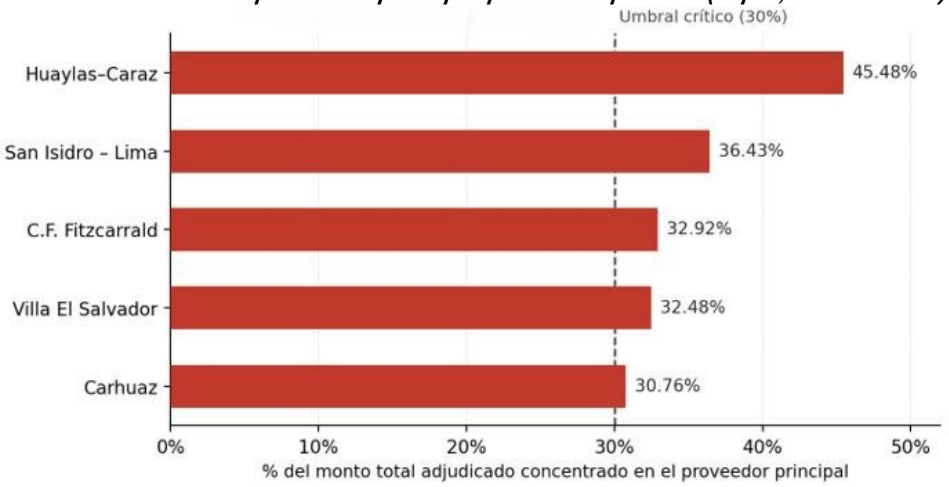
Municipalidad	Proveedor Principal	Monto Total Adjudicado (S/)	% Concentración
Huaylas–Caraz	EFF INGENIEROS E.I.R.L.	42,959,763.21	45.48%

Municipalidad	Proveedor Principal	Monto Total Adjudicado (S/)	% Concentración
San Isidro – Lima	RJV SOCIEDAD ANONIMA CERRADA	76,922,228.52	36.43%
C. F. Fitzcarrald	CODETECH S.A.C.	26,218,781.00	32.92%
Villa El Salvador	INVERSIONES LUMI E.I.R.L.	79,231,337.89	32.48%
Carhuaz	ALFET CONTRATISTAS GENERALES S.A.C.	21,093,638.20	30.76%

La Tabla 3 evidencia niveles extremos de concentración, destacando el caso de Huaylas-Caraz, donde un solo actor gestiona casi la mitad del presupuesto adjudicado de la entidad. Estos hallazgos confirman la existencia de mercados locales restringidos que debilitan los principios de libre concurrencia y competencia.

Figura 4

Concentración financiera en el proveedor principal por municipalidad (Top 5, 2022–2024)



Indicios de fraccionamiento contractual: Se detectaron casos de acumulación de contratos menores a 8 UIT hacia un mismo proveedor. En la Municipalidad Provincial de Casma, se identificó un proveedor con 11 contratos que sumaron S/ 388,368.62, mientras que en Ate-Vitarte un proveedor acumuló 9 contratos por un total de S/ 347,155.86.

Tabla 4

Proveedores con indicios de fraccionamiento por municipalidad

Municipalidad	Proveedor	Nº de Contrataciones Directas	Monto Total (S/)	Promedio por Contratación (S/)	Umbral Legal (UIT)
San Marcos	VARGAS SALAZAR JORGE JULIO	6	219,597.00	36,599.50	8 UIT
Casma	APESTEGUI ALEGRE OSCAR ROLANDO	11	388,368.62	35,306.24	8 UIT

AÑO 4. No. 3
ISSN:2992-7978

Ate–Vitarte	KURMI COLOR S.A.C.	9	347,155.86	38,572.87	8 UIT	

Riesgo relacional: Miembro Comité - Proveedor El análisis de recurrencias identificó que en la Municipalidad Distrital de San Marcos, el miembro de comité Eladio Yonny Romani Mautino registró 18 coincidencias con el proveedor Antúnez Celmi Florentino Antonio.

Tabla 5

Riesgo Relacional: Miembro Comité - Proveedor

Entidad	Miembro de Comité	Proveedor	Coincidencia
MUNICIPALIDAD DISTRITAL DE SAN MARCOS	ELADIO YONNY ROMANI MAUTINO	ANTUNEZ CELMI FLORENTINO ANTONIO	18
MUNICIPALIDAD DISTRITAL DE SAN MARCOS	ELADIO YONNY ROMANI MAUTINO	YAEV E.I.R.L.	17
MUNICIPALIDAD DISTRITAL DE ATE - VITARTE	CARLOS HUMBERTO VEGA ESPINOZA	H H H CONSTRUCCIONES Y SERVICIOS S.A.C.	11
MUNICIPALIDAD DISTRITAL DE ATE - VITARTE	CARLOS HUMBERTO VEGA ESPINOZA	FCL CONSTRUCCIONES Y SERVICIOS S.A.C.	8
MUNICIPALIDAD DISTRITAL DE CHORRILLOS	EDWIN RICARDO MUÑOZ URBINA	CONSTRUCTORA Y SERVICIOS VALDIVIA S.R.L.	5

Nivel de competencia. El análisis de la pluralidad de postores permitió cuantificar el nivel de concurrencia en los procesos de selección. Como se observa en la Tabla 6, existe una marcada predominancia de procesos con un solo postor, lo cual es especialmente crítico en la Municipalidad de San Marcos, donde el 85.5% de los casos (676 procesos) no tuvo competencia efectiva. Este hallazgo constituye un indicio de riesgo de direccionamiento, ya que la escasa participación limita la capacidad de la entidad para obtener mejores condiciones de precio y calidad.

Figura 5

Distribución de la pluralidad de postores por proceso, por municipalidad (2022–2024)

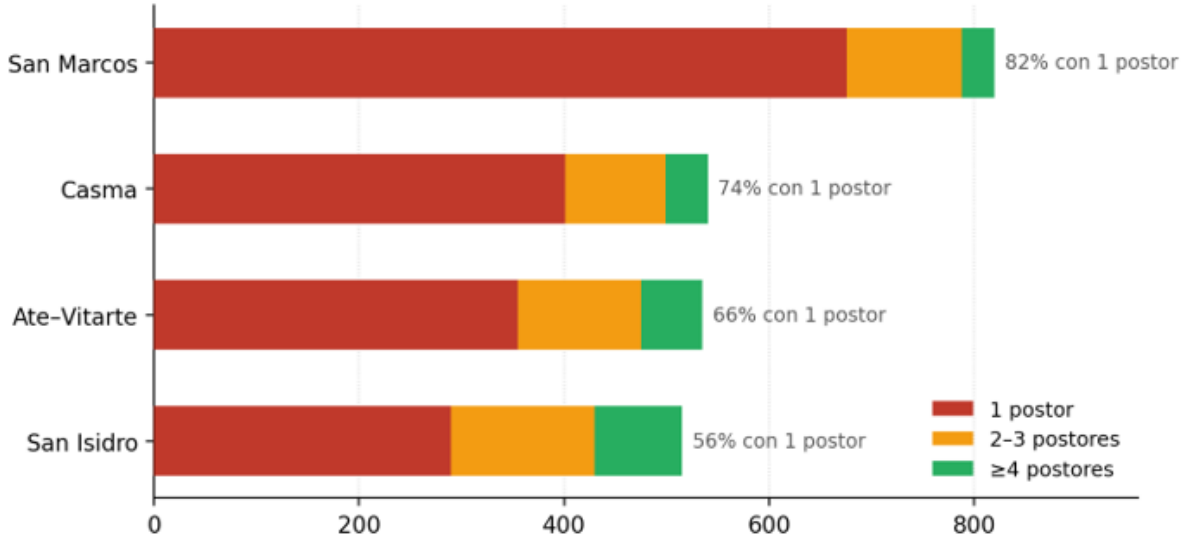


Tabla 6

Distribución del número de postores por proceso y municipalidad (2022–2024)

Municipalidad	1 postor	2–3 postores	≥4 postores
San Marcos	676	112	32
Casma	401	98	41
Ate–Vitarte	355	120	60
San Isidro	290	140	85

DISCUSIÓN

Los hallazgos empíricos confirman que la analítica de datos es una herramienta eficaz para identificar áreas de riesgo y mejorar la transparencia en la gestión pública local, en tal sentido, el éxito en la cuantificación de estos patrones depende directamente de la superación de la barrera técnica de la fragmentación de la información del OSCE mediante procesos ETL, lo que valida que la calidad del dato es el factor habilitador para la supervisión basada en evidencia.

En cuanto a la concentración de proveedores, los niveles extremos observados, como el 45.48% en la Municipalidad de Huaylas-Caraz (EFF Ingenieros E.I.R.L.) o el 36.43% en San Isidro, refuerzan la noción de mercados locales restringidos, concentración que genera una dependencia institucional crítica que debilita la libre competencia y es consistente con las advertencias de la Contraloría General de la República (2023,2024) sobre factores de riesgo en gobiernos locales.

El hallazgo de que en la Municipalidad de San Marcos, referida a la pluralidad de postores, donde el 85.5% de los procesos (676 casos) contaron con un único postor representa una vulneración directa al principio de pluralidad de postores consagrado en la Ley N.º 30225; asimismo, desde la perspectiva de la economía de la corrupción, esta baja competencia efectiva facilita adjudicaciones ineficientes y posibles sobrevaloraciones, al reducir la presión comparativa que normalmente disciplina los precios y las condiciones ofertadas por los proveedores.

En la Municipalidad de Casma, donde se identificó que un solo proveedor acumuló 11 contratos por un total de S/ 388,368.62, respalda la hipótesis de que esta práctica podría orientarse a eludir los procesos de selección formales, mediante el fraccionamiento contractual, lo que contraviene la prohibición expresa del Artículo 13 de la Ley N.º 30225 y coincide con lo señalado por Salazar y Angles (2018) respecto a la fragmentación del gasto como estrategia para evadir mecanismos de control.

Asimismo, las 18 coincidencias, referidas al riesgo relacional, identificadas entre el miembro del comité Eladio Yonny Romani Mautino y el proveedor Antúnez Celmi Florentino Antonio en la municipalidad de San Marcos representa una alerta analítica de posible atipicidad operacional que por sí misma no constituye una prueba de parcialidad o acto ilícito, pero señala un patrón relacional que los métodos de auditoría tradicionales suelen omitir y que requiere verificación administrativa, dado que la recurrencia transversal de estas relaciones en diversas municipalidades sugiere un riesgo de direccionamiento sistémico que los modelos analíticos basados en herramientas como el SQL logran capturar con precisión técnica.

Además, se reconoce como limitación principal la dependencia total de la calidad de los datos del Portal de Datos Abiertos del OSCE, dado que la fragmentación y heterogeneidad de la información de origen obliga a una preparación manual exhaustiva, lo que limita la inmediatez del monitoreo en tiempo real; además, los resultados se circunscriben a las 20 municipalidades analizadas durante el periodo 2022–2024, por lo que no pueden generalizarse a la totalidad del Estado peruano sin una expansión del modelo analítico.

Finalmente, debe considerarse el sesgo potencial derivado de la incompletitud e inconsistencia de origen en los registros de datos abiertos del OSCE, lo cual exige que toda alerta analítica sea contrastada con expedientes físicos antes de emitir un juicio definitivo

CONCLUSIONES

Se concluye que la aplicación de CRISP-DM y procesos ETL permite superar la fragmentación de los datos abiertos para identificar riesgos de manera sistemática, confirmando que la analítica de datos es eficaz para cuantificar patrones de concentración, fraccionamiento que afectan la transparencia local, riesgos relacionales, así como bajos niveles de competencia, precisando que por su carácter exploratorio, los hallazgos representan indicios de riesgo y no pruebas concluyentes de corrupción sin una validación externa.

En términos éticos y legales, el estudio se rige por el principio de transparencia proactiva, utilizando únicamente datos de carácter público, sin embargo, para el manejo de la información de los miembros de comités y representantes legales, se consideró la Ley N° 29733 (Ley de Protección de Datos Personales), asegurando que la generación de alertas de riesgo se limite al ejercicio del control social y la vigilancia ciudadana sobre el uso de recursos públicos, evitando la difusión de datos sensibles que no resulten estrictamente relevantes para la integridad de la contratación como es el caso de los documentos de identidad.

Finalmente, se recomienda al Organismo Supervisor de las Contrataciones del Estado (OSCE) la implementación de procesos ETL automatizados y tableros analíticos en tiempo real dentro de sus portales de datos abiertos, con el fin de fortalecer el monitoreo continuo, la auditoría preventiva y la fiscalización ciudadana basada en evidencia.

REFERENCIAS

Contraloría General de la República. (2023, 22 de mayo). *Corrupción e inconducta funcional habrían ocasionado pérdidas por S/ 24 268 millones en el 2023*. https://www.gob.pe/institucion/contraloria/noticias/912182-corrupcion-e-inconducta-funcional-habrian-ocasionado-perdidas-por-s-24-268-millones-en-el-2023

Contraloría General de la República. (2024, 15 de julio). *Contraloría identifica a las 321 entidades públicas con mayores riesgos de corrupción en sus contrataciones del Estado*. https://www.gob.pe/institucion/contraloria/noticias/988243-contraloria-identifica-a-las-321-entidades-publicas-con-mayores-riesgos-de-corrupcion-en-sus-contrataciones-del-estado

Felizzola, H., Gomez, C., Arrieta, N., Jerez, V., Erazo, Y., & Camacho, G. (2024). Enhancing transparency in public procurement: A data-driven analytics approach. *Information Systems, 125*, Article 102430. https://www.sciencedirect.com/science/article/abs/pii/S0306437924000887

García-Rodríguez, J., Pérez-Sánchez, M., & Gómez-López, R. (2025). Modelos de aprendizaje automático para clasificar el riesgo de corrupción en contratación pública. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (59), 1–20. https://scielo.pt/pdf/rist/n59/2183-0126-rist-59-3.pdf

Jorquera, M. (2019). *Compras públicas y Big Data: Investigación en Chile sobre índice de riesgo de corrupción.* Espacio Público. https://www.repository.fedesarrollo.org.co/handle/11445/3871

Ley de Contrataciones del Estado. (2014). Decreto Legislativo N.º 30225 y sus modificatorias. Diario Oficial El Peruano. https://www.gob.pe/institucion/osce/normatividad/214343-ley-n-30225

Muñoz-Cancino, R., & Ríos, S. A. (2025). Data-driven transparency: Machine learning and social network analysis for corruption detection in public procurement. *Procedia Computer Science, 270*, 1788–1795. https://www.sciencedirect.com/science/article/pii/S1877050925029722

Proética. (2021). *Análisis de la usabilidad de datos abiertos del OSCE durante la pandemia COVID-19.* Capítulo Peruano de Transparencia Internacional. https://www.proetica.org.pe/contenido/analisis-de-los-datos-abiertos-sobre-contrataciones-publicas-en-el-marco-de-la-emergencia-sanitaria-contra-el-covid-19/

RedelA. (2021). *Guía para la identificación de riesgos de corrupción en contratación pública utilizando la ciencia de datos*. https://ricg.org/wp-content/uploads/2021/12/Guia-para-la-identificacion-de-riesgos-de-corrupcion-en-CP-utilizando-la-ciencia-de-datos.pdf

Salazar, Z., & Angles, L. (2018). *La problemática de las contrataciones públicas en el Perú*. https://dialnet.unirioja.es/descarga/articulo/10274023.pdf