



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado
Facultad de Ingeniería de Sistemas e Informática
Unidad de Posgrado

**Modelo holístico de analítica avanzada para optimizar
la toma de decisiones organizacionales**

TESIS

Para optar el Grado Académico de Doctor en Ingeniería de
Sistemas e Informática

AUTOR

Juan Carlos MORALES AREVALO

ASESOR

Dr. Ciro RODRIGUEZ RODRIGUEZ

Lima, Perú

2026



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

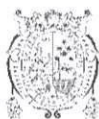
Morales, J. (2026). *Modelo holístico de analítica avanzada para optimizar la toma de decisiones organizacionales*. [Tesis de doctorado, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Unidad de Posgrado]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Juan Carlos Morales Arevalo
Tipo de documento de identidad	DNI
Número de documento de identidad	40007650
URL de ORCID	https://orcid.org/0000-0002-1834-6451
Datos de asesor	
Nombres y apellidos	Ciro Rodriguez Rodriguez
Tipo de documento de identidad	DNI
Número de documento de identidad	06020241
URL de ORCID	https://orcid.org/0000-0003-2112-1349
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Luis Soto Soto
Tipo de documento de identidad	DNI
Número de documento de identidad	22409532
Miembro del jurado 1	
Nombres y apellidos	Pedro Martin Lezama Gonzales
Tipo de documento de identidad	DNI
Número de documento de identidad	09656793
Miembro del jurado 2	
Nombres y apellidos	Ivan Carlo Petrlik Azabache

Tipo de documento de identidad	DNI
Número de documento de identidad	10140461
Datos de investigación	
Línea de investigación	C.0.4.1. Desarrollo de modelos, simulación y optimización de procesos

Grupo de investigación	No Aplica.
Agencia de financiamiento	Sin financiamiento
Ubicación geográfica de la investigación	Facultad de Ingeniería de Sistemas e Informática País: Perú Departamento: Lima Provincia: Lima Distrito: Lima Calle: Av. Amezaga Latitud: -12.05642315 Longitud: -77.08433269
Año o rango de años en que se realizó la investigación	2023 - 2025
URL de disciplinas OCDE	Ingeniería de sistemas y comunicaciones https://purl.org/pe-repo/ocde/ford#2.02.04



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
Universidad del Perú. Decana de América
Facultad de Ingeniería de Sistemas e Informática
Vicedecanato de Investigación y Posgrado
Unidad de Posgrado

**ACTA DE SUSTENTACIÓN DE TESIS PARA OPTAR EL GRADO ACADÉMICO DE
DOCTOR EN INGENIERÍA DE SISTEMAS E INFORMÁTICA**

A los veintisiete (27) días del mes de marzo de 2026, siendo las 11:00 am., se reunieron en el Auditorio Alfredo Celso Alva Bravo, el Jurado de Tesis conformado por los siguientes docentes:

Dr. Luis Soto Soto (Presidente)
Dr. Pedro Martín Lezama Gonzales (Miembro)
Dr. Ivan Carlo Petrlik Azabache (Miembro)
Dr. Ciro Rodríguez Rodríguez (Miembro Asesor)

Se inició la Sustentación invitando al candidato a Doctor **JUAN CARLOS MORALES AREVALO**, para que realice la exposición oral de la tesis para optar el Grado Académico de Doctor en Ingeniería de Sistemas e Informática, siendo la Tesis intitulada:

**“MODELO HOLÍSTICO DE ANALÍTICA AVANZADA PARA OPTIMIZAR LA TOMA
DE DECISIONES ORGANIZACIONALES”**

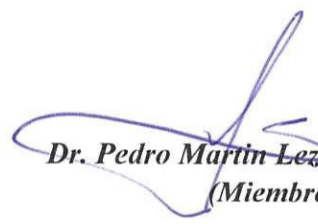
Concluida la exposición, los miembros del Jurado de Tesis procedieron a formular sus preguntas que fueron absueltas por el graduando; acto seguido se procedió a la evaluación correspondiente, habiendo obtenido la siguiente calificación:

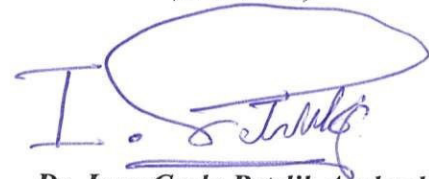
Diecisiete (17) *Muy bueno*

Por tanto, el presidente del Jurado, de acuerdo con el Reglamento General de Estudios de Posgrado, otorga al Magister **JUAN CARLOS MORALES AREVALO** el Grado Académico de Doctor en Ingeniería de Sistemas e Informática.

Siendo las *12:00* horas, el presidente del Jurado de Tesis, da por concluido el acto académico de Sustentación de Tesis.


Dr. Luis Soto Soto
(Presidente)


Dr. Pedro Martín Lezama Gonzales
(Miembro)


Dr. Ivan Carlo Petrlik Azabache
(Miembro)


Dr. Ciro Rodríguez Rodríguez
(Miembro Asesor)



Anexo 1

CERTIFICADO DE SIMILITUD

Yo, **Ciro Rodriguez Rodriguez** en mi condición de asesor acreditado con Dictamen N°000562-2024-UPG-VDIP-FISI/UNMSM de la tesis, cuyo título es “**Modelo Holístico de Analítica Avanzada para Optimizar la Toma de Decisiones Organizacionales**” presentado por el magíster **Juan Carlos Morales Arevalo** para optar el Grado Académico de Doctor en Ingeniería de Sistemas e Informática,

CERTIFICO que se ha cumplido con lo establecido en la Directiva de Originalidad y de Similitud de Trabajos Académicos, de Investigación y Producción Intelectual. Según la revisión, análisis y evaluación mediante la plataforma PIDS, el documento evaluado cuenta con el **9%** de similitud, nivel **PERMITIDO** para continuar con los trámites correspondientes y para su **publicación en el repositorio institucional**.

Se emite el presente certificado en cumplimiento de lo establecido en las normas vigentes, como uno de los requisitos para la obtención del Grado Académico de Doctor en Ingeniería de Sistemas e Informática correspondiente.

Lima, 18 de febrero de 2026

Firma del Asesor

Nombres y Apellidos del asesor: **Ciro Rodriguez Rodriguez**

DNI: 06020241

* Según corresponda: bachiller, magíster, doctor; título profesional o título de segunda especialidad

La vigencia del presente certificado corresponde a un período de noventa (90) días desde la fecha de su emisión; siempre que no haya observaciones en el texto durante el proceso de revisión por el jurado informante o surja alguna contingencia.

DEDICATORIA

A Dios, por darme la fuerza, la claridad y la constancia para llegar hasta el final de
este camino.

A mis hermanos Morales-Arevalo, por estar siempre ahí, con su apoyo, su paciencia
y ese acompañamiento incondicional que hizo más llevadero cada paso de este
proceso académico.

AGRADECIMIENTOS

Al doctor Ciro Rodríguez, mi asesor de tesis, por su orientación clara, honesta y generosa a lo largo de todo este proceso. Su confianza, franqueza y disposición para entablar un diálogo reflexivo han sido fundamentales para dar forma a este trabajo y para mi crecimiento como investigador.

También agradezco a los profesores del programa de doctorado sus altos estándares y sus valiosas aportaciones, que han fortalecido mi pensamiento crítico y mi desarrollo académico.

Por último, estoy muy agradecido a mis compañeros de promoción. Compartir este viaje con todos ustedes ha hecho que la experiencia del doctorado haya sido académicamente estimulante y personalmente significativa.

Índice General

RESUMEN.....	1
CAPÍTULO 1: INTRODUCCIÓN	3
1.1 Situación Problemática	3
1.2 Formulación del Problema	4
1.2.1 Problema General.....	4
1.2.2 Problemas específicos	4
1.3 Justificación Teórica.....	5
1.4 Justificación Práctica	5
1.5 Objetivos.....	5
1.5.1 Objetivo General	6
1.5.2 Objetivos Específicos.....	7
CAPÍTULO 2: MARCO TEÓRICO.....	8
2.1 Marco Epistemológico de la Investigación	8
2.1.1 Enfoque Pragmatista	8
2.1.2 Fundamentos de la Epistemología Adoptada	9
2.1.3 Aplicación en la Metodología	9
2.2 Antecedentes del Problema.....	9
2.2.1 Modelos de Madurez en el Análisis Avanzado de Datos (ADA).....	10
2.2.2 Aplicaciones y Factores en PYMES y Educación Superior.....	11
2.2.3 Impulsores y Retos Metodológicos	12
2.2.4 Calidad, Privacidad y Seguridad de los Datos	13
2.2.5 Modelos de Análisis de Datos y Toma de Decisiones	15
2.2.6 Metodologías de Diseño, Análisis y Gestión de Datos	17
2.2.7 Calidad, Seguridad y Metadatos.....	20
2.2.8 Visualización de Datos.....	21
2.3 Bases Teóricas	21
2.4 Glosario de términos.	24
CAPÍTULO 3: METODOLOGÍA.....	25
3.1 Tipo y Diseño de Investigación	26
3.1.1 Tipo de Investigación.....	26
3.1.2 Enfoque metodológico: Design Science Research (DSR)	27
3.1.3 Diseño de la investigación basado en DSR.....	27
3.1.4 Metodología para el Proceso de Toma de Decisiones.....	29
3.2 Propositiones.....	31
3.2.1 Proposición general	31

3.2.2	Proposiciones específicas.....	31
3.3	Dimensiones del modelo y criterios de evaluación.....	32
3.3.1	Identificación de dimensiones del modelo	32
3.3.2	Operacionalización de criterios de evaluación.....	33
3.4	Matriz de Consistencia.....	36
3.5	Unidad de análisis.....	39
3.6	Población y muestra	39
3.6.1	Población de estudio	39
3.6.2	Tamaño de la muestra	39
3.6.3	Selección de muestra.....	40
3.7	Técnicas de recolección de Datos	40
3.7.1	Revisión sistemática de la literatura.....	40
3.7.2	Juicio de expertos.....	40
3.7.3	Caso aplicado	43
3.8	Procesamiento y análisis de datos	43
3.8.1	Análisis de contenido de la literatura científica	43
3.8.2	Análisis de los resultados del juicio de expertos.....	43
3.8.3	Análisis del caso aplicado mediante datos abiertos.....	44
CAPÍTULO 4: DESARROLLO DEL MODELO HOLÍSTICO DE ANALÍTICA		
AVANZADA.....		45
4.1	Estructuración de las fases del modelo holístico de analítica avanzada para	
	establecer la arquitectura del artefacto analítico.....	45
4.1.1	Identificación del problema y motivación (Fase DSR 1)	45
4.1.2	Definir los objetivos de una solución (Fase DSR 2)	46
4.2	Diseño del Modelo Holístico de Analítica Avanzada mediante la	
	descomposición funcional de su arquitectura, especificando componentes y flujos de	
	integración.....	49
4.2.1	Diseño y desarrollo: artefacto (Fase DSR 3).....	49
4.2.2	Demostración (Fase DSR 4).....	55
4.3	Validación del Modelo Holístico de Analítica Avanzada mediante el juicio de	
	expertos y un caso aplicado	56
4.3.1	Evaluación (Fase DSR 5)	56
4.3.2	Comunicación (Fase DSR 6).....	87
CAPÍTULO 5: RESULTADOS Y DISCUSIÓN		89
5.1	Resultados	89
5.1.1	Resultados del Juicio de Expertos.....	89

5.1.2	Resultados del Caso Aplicado.....	90
5.1.3	Evaluación del Cumplimiento de las Proposiciones	91
5.1.4	Triangulación de Evidencia.....	93
5.1.5	Argumento de Optimización de la Toma de Decisiones	94
5.2	Discusión.....	95
	CONCLUSIONES.....	98
	RECOMENDACIONES.....	101
	REFERENCIAS BIBLIOGRÁFICAS	102
	ANEXO A: Código en Python Fase Estructura de Contexto Organizacional.....	112
	ANEXO B: Código en Python Fase Limpieza y Calidad de Datos.....	113
	ANEXO C: Código en Python Fase Integración de datos	118
	ANEXO D: Código en Python Fase Analítica Avanzada	126

Lista de Tablas

Tabla 1	Clasificación de los tipos de decisión	4
Tabla 2	Matriz de Operacionalización de criterios de evaluación del modelo	33
Tabla 3	Matriz de Consistencia	36
Tabla 4	Preguntas del instrumento de validación y criterios evaluados	42
Tabla 5	Estructura conceptual del Modelo Holístico de Analítica Avanzada	48
Tabla 6	Contexto Organizacional	50
Tabla 7	Limpieza y calidad de los datos	50
Tabla 8	Integración de Datos	51
Tabla 9	Capacidades de Analítica Avanzada	52
Tabla 10	Visualización de Datos	53
Tabla 11	Alineación del modelo holístico y el modelo de toma de decisiones	55
Tabla 12	Resultados de la validación del modelo según el criterio de los expertos.....	57
Tabla 13	Nivel de aceptación del modelo por criterios	58
Tabla 14	Perfil de calidad del dataset antes del proceso de limpieza	61
Tabla 15	Resultados de las reglas de calidad (antes de la limpieza).....	61
Tabla 16	Detección de duplicados y consolidación analítica	62
Tabla 17	Perfil de calidad del dataset después del proceso de limpieza.....	62
Tabla 18	Distribución y control de valores extremos (outliers)	63
Tabla 19	KPIs del proceso de integración de datos	63
Tabla 20	Resultados de la validación semántica del dataset integrado.....	64
Tabla 21	Resumen de dominios y catálogos integrados.....	65
Tabla 22	Estructura del dataset integrado (diccionario resumido).....	65
Tabla 23	Indicadores descriptivos globales del sistema de atenciones	66
Tabla 24	Tendencia Mensual año 2020	67
Tabla 25	Concentración de atenciones.....	68
Tabla 26	Métricas del modelo predictivo.....	69
Tabla 27	Comparación Real vs Predicho (casos con mayor error)	70
Tabla 28	Configuración analítica del modelo	71
Tabla 29	Recomendaciones prescriptivas (Top 30 – Prioridad ALTA).....	72
Tabla 30	Resumen de prioridades prescriptivas	79
Tabla 31	Escenarios priorizados según demanda estimada de atenciones	79
Tabla 32	Evaluación del cumplimiento de las proposiciones del MHAA	92
Tabla 33	Matriz de Evidencia Empírica de la Prueba de Concepto	94
Tabla 34	Evaluación comparativa estructural de modelos de analítica respecto a la optimización de la TD.....	97

Lista de Figuras

Figura 1 Niveles de sofisticación de la analítica y ventaja competitiva.....	16
Figura 2 Fases del modelo CRISP-DM	18
Figura 3 Ciclo de vida de la metodología Kimball	19
Figura 4 DSRM Process Model.....	27
Figura 5 Modelo de proceso de toma de decisiones en tres fases.....	30
Figura 6 Arquitectura Estructurada del Modelo Holístico de Analítica Avanzada	49
Figura 7 Modelo Holístico de Analítica Avanzada.....	54
Figura 8 Evolución mensual del total de atenciones de salud (2020).....	68
Figura 9 Dashboard Analítica Descriptiva	83
Figura 10 Dashboard Analítica Predictiva	85
Figura 11 Dashboard Analítica Prescriptiva.....	87
Figura 12 Resultados del juicio de expertos (pertinencia, viabilidad y utilidad)	90

RESUMEN

En las organizaciones actuales, la toma de decisiones es cada vez más compleja. Esto se debe principalmente a que los datos suelen estar dispersos, presentan problemas de calidad y no siempre existen arquitecturas analíticas integradas que permitan transformar esa información en conocimiento accionable. Aunque la analítica avanzada y la toma de decisiones han sido ampliamente estudiadas, persiste una limitada articulación estructural entre ambas dentro de un mismo marco arquitectónico. Esta tesis doctoral aborda ese vacío mediante el diseño y validación de un Modelo Holístico de Analítica Avanzada orientado a optimizar la toma de decisiones organizacionales.

Desde una perspectiva pragmática, la investigación se inscribe en la metodología Design Science Research (DSR), privilegiando el diseño, la construcción y la evaluación de un artefacto analítico por encima de la mera contrastación de hipótesis estadísticas. En relación con el Objetivo Específico 1, las fases del modelo fueron configuradas mediante la integración sistemática de marcos teóricos y el examen crítico de la literatura especializada, con lo cual quedó delineada su arquitectura conceptual. En coherencia con el Objetivo Específico 2, el Modelo Holístico de Analítica Avanzada se concibió a partir de la descomposición funcional de sus componentes y de la articulación de flujos de integración alineados con el proceso clásico de toma de decisiones —inteligencia, diseño y elección—.

Respecto al Objetivo Específico 3, la validación se articuló mediante un juicio estructurado de expertos orientado a valorar pertinencia, utilidad y viabilidad. Con posterioridad, se implementó un caso aplicado con datos abiertos, incorporando técnicas de analítica avanzada en Python, lo que permitió constatar eficacia operativa, incremento en la confiabilidad de los datos, mayor coherencia analítica y precisión predictiva; tales resultados posibilitaron sustentar decisiones más consistentes y oportunas. En conjunto, los hallazgos permiten sostener el cumplimiento del objetivo general: desarrollar un modelo holístico de analítica avanzada con capacidad para optimizar la toma de decisiones organizacionales.

Palabras clave: *Analítica avanzada; Toma de decisiones organizacionales; Modelo holístico; Investigación en Ciencias del Diseño; Toma de decisiones basada en datos.*

ABSTRACT

In contemporary organizations, decision-making is becoming increasingly complex. This is primarily due to the fact that data are often dispersed, exhibit quality issues, and lack integrated analytical architectures capable of transforming information into actionable knowledge. Although advanced analytics and decision-making have been widely studied, a limited structural articulation between them persists within a single architectural framework. This doctoral dissertation addresses that gap through the design and validation of a Holistic Advanced Analytics Model aimed at optimizing organizational decision-making.

From a pragmatic perspective, the research adopts the Design Science Research (DSR) methodology, prioritizing the design, construction, and evaluation of an analytical artifact over the statistical testing of hypotheses. In relation to Specific Objective 1, the phases of the model were structured through the systematic integration of theoretical frameworks and the analysis of specialized literature, establishing its conceptual architecture. In correspondence with Specific Objective 2, the Holistic Advanced Analytics Model was designed through the functional decomposition of its components and the articulation of integration flows aligned with the classical decision-making process: intelligence, design, and choice.

With respect to Specific Objective 3, validation was conducted through a structured expert judgment process, assessing relevance, usefulness, and feasibility. Subsequently, an applied case study using open data was developed, employing advanced analytics techniques in Python, demonstrating operational effectiveness, improvement in data reliability, analytical coherence, and predictive accuracy, which enabled the support of more consistent and timely decisions. Overall, the results confirm the fulfillment of the general objective of developing a holistic advanced analytics model capable of optimizing organizational decision-making.

Keywords: *Advanced analytics; Organizational decision-making; Holistic model; Design Science Research; Data-driven decision-making*

CAPÍTULO 1: INTRODUCCIÓN

1.1 Situación Problemática

Hoy en día una decisión bien tomada puede impulsar el crecimiento y la rentabilidad de una organización, mientras que una decisión equivocada puede tener consecuencias negativas significativas, por consiguiente, Quituizaca et al. (2022) afirman que una gestión adecuada del proceso decisional resulta fundamental para el desarrollo y la sostenibilidad de una empresa, ya que decisiones deficientes, sostenidas en el tiempo, pueden derivar en escenarios de fracaso e incluso en su desaparición; en la misma línea, Tuanama et al. (2019) sostienen que las decisiones gerenciales, desde las primeras etapas de la organización, cumplen un rol determinante en su desempeño, al influir directamente en el control del negocio y en la probabilidad de éxito o declive, dado que una decisión inadecuada incrementa significativamente el riesgo de fracaso organizacional; basándose en su análisis de 400 decisiones estratégicas tomadas por altos directivos en áreas como productos y servicios, precios y mercados, política de personal, adquisición de tecnología y reorganización estratégica, Nutt (2002) estima que dos tercios de todas las decisiones se basan en tácticas cuestionables o propensas al fracaso; del mismo modo, un estudio de Business Wire (2022) detectó que el 82% de las empresas toman decisiones basadas en información desactualizada, lo que lleva a decisiones incorrectas y pérdida de ingreso.

En la misma línea, Asikhia y Mba (2021), destacan cómo decisiones poco fundamentadas pueden desviar la dirección estratégica y generar impactos adversos en la competitividad organizacional. En la Tabla 1, podemos observar que la toma de decisiones bajo incertidumbre está relacionada directamente con la información, la cual carece de claridad, exactitud y/o agilidad, según Kuchta et al. (2023), la incertidumbre se define como la falta de conocimiento en determinadas áreas que son cruciales para la comprensión y gestión de un proyecto.

Tabla 1***Clasificación de los tipos de decisión***

Tipo de Decisión	Información	Riesgo
Condiciones de Certeza	Clara, exacta y completa	Bajo
Condiciones de Riesgo	Con probabilidad de ocurrencia	Medio
Condiciones de Incertidumbre	Muy escasa, nula, inexacta, inoportuna	Alto

Nota: Cuando las decisiones se toman bajo condiciones de incertidumbre, el decisor no dispone de información suficiente sobre los posibles resultados de los distintos estados de la naturaleza, o bien el acceso a dicha información resulta limitado o costoso. En este contexto, el proceso decisional se ve fuertemente influenciado por las características personales del decisor, lo que incrementa significativamente el nivel de riesgo asociado a la decisión. *Fuente.* Tomado de Peñaloza (2010)

1.2 Formulación del Problema

En un contexto empresarial cada vez más complejo y competitivo, las organizaciones enfrentan el desafío de procesar, analizar y entender la enorme cantidad de datos disponibles para tomar decisiones informadas y estratégicas (Davenport y Harris, 2007).

1.2.1 Problema General

¿Cómo el desarrollo de un modelo holístico de analítica avanzada optimiza la toma de decisiones organizacionales?

1.2.2 Problemas específicos

- PE1: ¿Cómo la estructuración de las fases del Modelo Holístico de Analítica Avanzada, mediante la integración sistemática de marcos teóricos y el análisis sistemático de la literatura, establece la arquitectura del artefacto analítico?

- PE2: ¿Cómo el diseño del Modelo Holístico de Analítica Avanzada mediante la descomposición funcional de su arquitectura, especificando componentes y flujos de integración, se articula con el proceso de toma de decisiones organizacionales?
- PE3: ¿Cómo la validación del Modelo Holístico de Analítica Avanzada, mediante el juicio de expertos, evidencia niveles altos de pertinencia, utilidad y viabilidad, y cómo, mediante un caso aplicado, demuestra su eficacia en la optimización de la toma de decisiones organizacionales?

1.3 Justificación Teórica

Hasta ahora, se han llevado a cabo muchos estudios tanto sobre la toma de decisiones como sobre la analítica avanzada, pero casi siempre de forma separada. Lo que aún falta, y resulta clave, es profundizar en cómo se relacionan entre sí, especialmente en el contexto de las decisiones dentro de las organizaciones. Davenport y Harris (2007) señalan que, al incorporar la analítica avanzada en sus procesos de decisión, las empresas pueden volverse mucho más competitivas. Esto pone sobre la mesa la necesidad de explorar con mayor detalle esa conexión; por consiguiente, este estudio tiene como objetivo contribuir a la literatura examinando cómo influye el análisis avanzado en la toma de decisiones organizativas.

1.4 Justificación Práctica

La presente investigación sigue un diseño no experimental y propone un marco integrado que utiliza análisis avanzados para evaluar la toma de decisiones en las organizaciones. Como señalan Shmueli et al. (2017), el uso integrado de análisis avanzados puede mejorar la previsión y la eficacia de la toma de decisiones, lo que respalda la relevancia práctica de este enfoque.

1.5 Objetivos

Los objetivos de este estudio se basan en el reconocimiento de que muchas

organizaciones líderes ya confían en análisis avanzados para transformar los datos en conocimiento útil, mejorar la toma de decisiones y optimizar los procesos empresariales clave en distintas áreas de la organización (Davenport y Harris, 2017).

Razón por la cual el desarrollo de un modelo holístico de analítica avanzada requiere, como primer paso, definir con claridad sus fases y ordenar sus componentes esenciales. Según DAMA International (2017), modelar implica identificar y analizar las necesidades de información para representarlas con precisión a través de un modelo, que sirva como base estructural para organizar los elementos del fenómeno organizacional.

En la misma línea, Reis y Housley (2022) entienden el modelado como el proceso de estructurar y organizar datos de acuerdo con las necesidades del negocio y de quienes los usan, lo que justifica la importancia de una etapa inicial dedicada a dicha organización previa al diseño del modelo.

Es importante recordar que la validación se sustenta en los fundamentos del enfoque Design Science Research (DSR), ya que todo artefacto debe ser evaluado en función de su adecuación al contexto organizacional, de su capacidad para resolver el problema y de su viabilidad de aplicación (Hevner et al., 2004; March & Smith, 1995; Peffers et al., 2007).

En resumen, Peffers et al. (2007) resaltan la importancia de evaluar en qué medida el artefacto contribuye a la solución del problema planteado, en función de aspectos como la utilidad y la eficacia; razón por la cual es necesaria la inclusión de un caso práctico que permita demostrar de manera concreta cómo el modelo puede ayudar a optimizar los procesos de toma de decisiones en una organización.

1.5.1 Objetivo General

Desarrollar un modelo holístico de analítica avanzada para optimizar la toma de decisiones organizacionales.

1.5.2 Objetivos Específicos

- OE1: Estructurar las fases del modelo holístico de analítica avanzada mediante la integración sistemática de marcos teóricos y el análisis sistemático de la literatura, para establecer la arquitectura del artefacto analítico.
- OE2: Diseñar el Modelo Holístico de Analítica Avanzada mediante la descomposición funcional de su arquitectura, especificando componentes y flujos de integración para su articulación con el proceso de toma de decisiones organizacionales.
- OE3: Validar el modelo holístico de analítica avanzada mediante el juicio de expertos para evidenciar altos niveles de pertinencia, utilidad y viabilidad, y mediante un caso aplicado para demostrar su eficacia en la optimización de la toma de decisiones organizacionales.

CAPÍTULO 2: MARCO TEÓRICO

El capítulo establece conceptos clave de la analítica avanzada que sustentan la investigación y orientan su aplicación. Berge et al. (2023) destacan que un marco integral, capaz de abarcar tanto datos estructurados como no estructurados, ofrece una visión más completa para apoyar la toma de decisiones.

2.1 Marco Epistemológico de la Investigación

La tesis se apoya en un enfoque pragmatista que permite combinar perspectivas técnicas y humanas sobre la analítica y su influencia en la toma de decisiones. Este posicionamiento atiende fenómenos complejos en los que convergen indicadores tecnológicos con factores contextuales, como la percepción de los usuarios, la madurez tecnológica y las dificultades de implementación en entornos reales (Morgan, 2007).

2.1.1 Enfoque Pragmatista

El pragmatismo pone el acento en la utilidad del conocimiento y en su capacidad para resolver problemas en contextos específicos (Creswell, 2014). No obliga a elegir de manera rígida entre métodos cuantitativos y cualitativos, sino a adaptar la estrategia metodológica al problema. Resulta idóneo para construir un modelo integral que incorpore dimensiones técnicas, organizacionales y humanas. Como indican Tashakkori y Teddlie (2010), el pragmatismo favorece el uso complementario de métodos. Aquí, técnicas cuantitativas permiten evaluar la eficacia del modelo en un caso práctico, mientras que la evidencia cualitativa, basada en juicio de expertos, valida su utilidad, relevancia y viabilidad.

Tashakkori y Teddlie (2010) apuntan que el pragmatismo promueve el uso de diferentes metodologías sin quedarse con una sola forma de generar conocimiento. En esta investigación, se emplean técnicas cuantitativas para evaluar la eficacia del modelo en un caso práctico, mientras que el análisis cualitativo se usa para validar su utilidad, relevancia y viabilidad, a partir del juicio de expertos.

2.1.2 Fundamentos de la Epistemología Adoptada

El modelo integra tres perspectivas:

La dimensión objetiva asume que los sistemas analíticos trabajan con datos medibles y replicables, lo que habilita evaluar efectividad con atención a precisión algorítmica y mitigación de sesgos (Bhattacharjee, 2012).

La dimensión interpretativa y contextual reconoce que el valor de la analítica depende también de cultura organizacional, disposición al cambio y competencias del equipo, por lo que se requiere interpretar su influencia en la adopción del modelo (Saunders et al., 2019).

La dimensión pragmática se relaciona con la aplicación de la investigación en los desafíos del mundo real, ya que tiene como objetivo desarrollar una herramienta sólida y útil, adecuada a las necesidades específicas del contexto y capaz de convertirse en un recurso práctico (Morgan, 2007).

2.1.3 Aplicación en la Metodología

Esta investigación se asienta en el pragmatismo, donde el valor del conocimiento se mide por su utilidad y no solo por su coherencia teórica. Como bien señala Van Aken (2004), en las ciencias de diseño el objetivo es generar "reglas tecnológicas" que resuelvan problemas de gestión reales. Por eso, el enfoque mixto de este estudio no es casual; es una necesidad para evaluar el desempeño del modelo desde distintos ángulos. Lo cierto es que, al seguir la lógica de Simon (1996) sobre las ciencias de lo artificial, lo que importa aquí es la eficacia del artefacto. En lugar de buscar verdades absolutas, buscamos validar si el modelo holístico realmente funciona como un puente útil para optimizar la toma de decisiones en la práctica.

2.2 Antecedentes del Problema

En este punto, presentaremos las investigaciones revisadas, ordenadas por

subtemas:

2.2.1 Modelos de Madurez en el Análisis Avanzado de Datos (ADA)

En años recientes, los modelos de madurez en el análisis avanzado de datos han cobrado relevancia, ya que ayudan a comprender cómo se incorporan las capacidades analíticas al desempeño general de una organización. En esta línea, Korsten y su equipo (2022), en su estudio ADA-CMM: A capability maturity model for advanced data analytics, investigan de qué manera las habilidades asociadas con el análisis avanzado de datos (ADA) impactan los resultados del negocio. El objetivo principal de su investigación es identificar qué tan directamente se relaciona el nivel de madurez analítica con el valor que generan estas iniciativas, así como con el efecto que tienen sobre el desempeño general de la organización. Para lo cual, desarrollan un Modelo de Madurez de Capacidad de Análisis Avanzado de Datos (ADA-CMM), el cual desarrollaron con base en un estudio Delphi con expertos en el tema. Este modelo se organiza en seis grandes áreas de capacidad y 24 subcapacidades, que permiten mapear de forma clara y estructurada qué tan madura es una organización en cuanto a análisis de datos. En cuanto al enfoque metodológico, el estudio combina métodos cualitativos y cuantitativos: por un lado, el método Delphi; y, por otro, un análisis estadístico basado en ecuaciones estructurales (PLS-SEM), que permite validar la relación entre la madurez analítica, la creación de valor y el rendimiento empresarial.

Dentro de este panorama, la forma en que se asignan puntajes en estos modelos de madurez se vuelve clave para poder evaluar con objetividad qué tan desarrolladas están las capacidades analíticas de una empresa. En este sentido, Grossman (2018), en su trabajo A framework for evaluating the analytic maturity of an organization, se pregunta cómo el nivel de madurez en ADA condiciona el valor y los resultados de estas iniciativas. Su propuesta se basa en el concepto de "madurez analítica", que mide mediante un Analytic Maturity Level (AML) dividido en cinco niveles. A mayor nivel, mayor es la probabilidad de que los modelos analíticos sean válidos, aplicables y alineados con los objetivos estratégicos del negocio. Para estructurar este enfoque, retoma el conocido Capability Maturity Model (CMM) y desarrolla el Modelo de Madurez de Procesos Analíticos (APMM), que incluye áreas clave y prácticas

específicas para evaluar el nivel de madurez analítica de forma sistemática.

Desde otra perspectiva, Chen y Nath (2018), en el estudio *Business analytics maturity in organizations: Examining the relationships between managerial IT perception, business analytics maturity and success*, examinan cómo se relacionan la madurez del Business Analytics (BA), la percepción que tienen los directivos sobre las tecnologías de la información (TI) y el éxito organizacional. Su análisis parte de una visión sociotécnica, es decir, consideran que el grado de madurez en BA no depende solo de la tecnología que tenga la empresa, sino también de cómo se articula con la gestión y la estrategia. Con un enfoque cuantitativo a través de encuestas, midieron tanto la madurez y el éxito del BA como la percepción del impacto estratégico de las TI; dichos resultados respaldan la idea de que una mirada integral “que combine lo tecnológico con lo organizacional” es clave para entender la efectividad del análisis de datos en las empresas.

Finalmente, Belfo y Andreica (2018) plantean una metodología más integral para implementar soluciones de Business Intelligence y Analytics, con énfasis en cómo obtener conocimiento útil a partir de bases de datos. En su investigación, exploran el uso de algoritmos de equivalencia y representación canónica dentro de proyectos de Knowledge Discovery in Databases (KDD). Su objetivo es obtener resultados más enriquecidos en comparación con los métodos tradicionales. Se concentran en mejorar la calidad de las reglas de asociación generadas por estos algoritmos, lo que permite una representación más robusta del conocimiento obtenido. Esta propuesta metodológica se presenta como una alternativa que mejora los resultados del análisis y los hace más precisos y útiles para la toma de decisiones.

2.2.2 Aplicaciones y Factores en PYMES y Educación Superior

Baijens et al. (2022), en su investigación "Data analytics and SMEs: How maturity improves performance," analizan la madurez en el uso del análisis de datos en pequeñas y medianas empresas (PYMES) y su impacto en el desempeño organizacional. El problema central de esta investigación se relaciona con la escasa evidencia empírica sobre cómo los diferentes niveles de madurez analítica contribuyen

a la creación de valor y al rendimiento empresarial. El marco teórico de este estudio incluye una revisión de modelos consolidados como el Analytics Maturity Quotient (AMQ), el modelo DELTA Plus, el Data Analytics Maturity Indicator (DAMI) y la propuesta metodológica de TDWI. Estos enfoques ayudan a identificar dominios clave, factores críticos de éxito y las etapas de evolución que puede atravesar una organización en su camino hacia una analítica más avanzada y madura. En cuanto al método, la investigación se basa en una revisión sistemática de la literatura, complementada con una fase experimental respaldada por pruebas diagnósticas en línea. Los hallazgos destacan la urgente necesidad de contar con modelos de madurez más claros, replicables y respaldados por una base metodológica sólida.

Cardoso y Su (2022) analizaron un tema distinto, pero relacionado: la ausencia de un modelo integral para evaluar el nivel de madurez de Business Intelligence and Analytics (BIA) en las Instituciones de Educación Superior (IES). El propósito de su estudio es diseñar y validar un modelo de madurez que permita diagnosticar el nivel de madurez de BIA en este tipo de organizaciones y, a partir de ello, orientar procesos de mejora. El diseño de la investigación se fundamenta en el enfoque de Investigación en Ciencias del Diseño, siguiendo de manera sistemática las etapas de identificación del problema, diseño y desarrollo del artefacto, demostración de su aplicación, evaluación de su utilidad y comunicación de los resultados. La metodología integra principios del Design Science Research con métodos cualitativos, lo que permite construir un modelo de madurez de BIA específicamente adaptado al contexto de las IES y sustentado en criterios conceptuales y organizacionales.

2.2.3 Impulsores y Retos Metodológicos

Saravanabhavan et al. (2022), en el estudio Understanding the drivers of Business Analytics maturity for Indian firms, examinan los factores que influyen en la madurez del Business Analytics en empresas de la India, partiendo de un marco conceptual sustentado en el enfoque Tecnológico-Organizacional-Ambiental (TOE). A partir de este marco, los autores incorporan subfactores adaptados al contexto local, entre los que destaca el denominado “impulsor del catalizador analítico”, diseñado específicamente para reflejar las condiciones del entorno empresarial indio. La

investigación emplea un cuestionario distribuido en línea mediante Google Forms, estructurado en tres secciones: perfil organizacional, percepción sobre los impulsores de los modelos de madurez analítica y beneficios percibidos de su adopción. Los datos obtenidos se procesan con el software ADANCO, complementándose con estadísticas descriptivas y análisis de los resultados, lo que permite examinar las relaciones entre los impulsores identificados y la madurez analítica en el contexto estudiado.

Lemsa (2021), en la investigación titulada Retos del desarrollo de un modelo de madurez analítica avanzada, aborda como problema central la limitada claridad metodológica presente en el diseño de diversos modelos de evaluación de madurez analítica. El autor señala que, en numerosos casos, los factores, dimensiones y criterios empleados para determinar el nivel de madurez del ecosistema analítico organizacional no se documentan de forma suficiente, lo que dificulta su validación empírica y su aplicación en otros contextos. Este estudio parte de una revisión de modelos consolidados como el Analytics Maturity Quotient (AMQ), el modelo DELTA Plus, el Data Analytics Maturity Indicator (DAMI) y la metodología propuesta por TDWI; estos marcos ayudan a identificar dominios clave, factores que influyen en el éxito y las etapas que suelen recorrer las organizaciones a medida que avanzan en el uso estratégico de la analítica. La investigación se basa en una revisión sistemática de la literatura especializada, acompañada de una fase experimental con pruebas diagnósticas en línea; a partir de este enfoque, se identifican áreas de oportunidad y retos comunes en los procesos de madurez analítica, cuyos resultados subrayan la urgencia de contar con modelos más claros y replicables, respaldados por un sólido respaldo metodológico.

2.2.4 Calidad, Privacidad y Seguridad de los Datos

Según Liu et al. (2023), en ambientes de Big Data, se presentan desafíos significativos relacionados con integridad, consistencia, precisión, seguridad, y privacidad. Estos desafíos empeoran debido a la necesidad de la detección de valores atípicos y la gestión de una creciente complejidad en los datos.

Según Demirdöğen et al. (2021), existen barreras relacionadas con la

colaboración, comunicación, acceso a la información, y transparencia, derivadas de una confianza limitada, disputas por propiedad intelectual, pérdida de datos, y responsabilidades poco claras. Para superarlas, proponen entornos comunes de datos (CDE) y contratos de proyecto integrados (IPD), en lugar de modelos tradicionales como el DBB, que limitan la colaboración.

Por otro lado, el enfoque de toma de decisiones basada en datos (Data-Driven Decision Making) fomenta el uso sistemático de información en todo el proceso decisional. Según Jeyaprabha et al. (2025), es fundamental en organizaciones con grandes volúmenes de datos, ya que mejora la calidad de las decisiones a través de evidencia cuantitativa. Borges et al. (2021) resaltan su importancia en marketing y planificación estratégica, ya que permite el análisis serio de escenarios y alternativas.

Por otro lado, Vásquez et al. (2021) llaman la atención sobre los desafíos particulares que enfrentan las micro, pequeñas y medianas empresas (MIPYMES) en cuanto a la recopilación y la fiabilidad de datos sobre prácticas ambientales. En muchos casos, estos datos son escasos, poco documentados o no se les da un seguimiento adecuado, lo que dificulta su uso en procesos analíticos o en la toma de decisiones orientadas a la sostenibilidad. En muchos casos, la información sobre prácticas sustentables es escasa, poco documentada y con escaso seguimiento, lo que dificulta su uso en procesos de análisis o de toma de decisiones orientados a la sostenibilidad.

Esta preocupación se refuerza con lo señalado por Belfo y Andreica (2018), quienes destacan que la efectividad de los algoritmos y técnicas de representación depende directamente de la calidad e integridad de los datos, ya que información inexacta o incompleta puede generar resultados sesgados, además de dificultar la selección de algoritmos adecuados en proyectos de Knowledge Discovery in Databases (KDD).

Finalmente, Hindle y Vidgen (2018) destacan que compartir datos entre organizaciones no es tarea sencilla, ya que existen barreras importantes relacionadas con la privacidad, las leyes de protección de datos y la renuencia a divulgar

información sensible. A esto se suman desafíos éticos que tienen que ver con la equidad, la no discriminación y la protección de la confidencialidad de las personas involucradas.

2.2.5 Modelos de Análisis de Datos y Toma de Decisiones

Los modelos de toma de decisiones organizacionales han ido evolucionando para responder a escenarios marcados por la complejidad y la incertidumbre, especialmente en contextos críticos. En este sentido, Khan et al. (2025) proponen un modelo basado en Conjuntos Difusos Continuos Fraccionarios (FCFS), el cual permite representar la incertidumbre de forma más flexible y sensible mediante el uso de funciones continuas, en lugar de valores discretos. Este enfoque amplía la capacidad de análisis y mejora la precisión en la evaluación de alternativas, como ocurre en la selección de robots industriales, donde intervienen múltiples criterios técnicos que deben ser ponderados de manera simultánea.

Desde una perspectiva organizacional distinta, Romiti et al. (2025) analizan el proceso de toma de decisiones en organizaciones del sector salud durante situaciones de crisis e identifican tres enfoques predominantes. El primero corresponde al modelo de CEO, caracterizado por una toma de decisiones centralizada y autocrática; el segundo, al modelo de CEO-asesor, en el que se incorpora la participación de asesores clave; y el tercero, al modelo del Top Management Team (TMT), basado en un esquema de gestión colaborativa. Estos modelos evidencian cómo los líderes ajustan sus estilos de decisión en función del nivel de presión e incertidumbre, optando por mayores grados de centralización o colaboración según las circunstancias.

Por su parte, Chacoma y Billoni (2025) estudiaron la complejidad emergente en la toma de decisiones de jugadores de ajedrez y evidencian que la precisión de las decisiones varía en función de la competitividad y la criticidad de las posiciones. Los resultados resaltan la importancia de adoptar estrategias adaptativas para mejorar el desempeño en contextos altamente competitivos, donde la optimización de métricas de decisión resulta determinante.

En conjunto, estos estudios ponen de manifiesto la necesidad de adaptar los modelos de decisión al contexto organizacional y a la naturaleza específica del problema, ya sea de carácter técnico o estratégico. En esta línea, el modelo de la Cadena de Valor de la Analítica aporta una visión de madurez sobre cómo las organizaciones transforman los datos en decisiones. Davenport y Harris (2007) identifican cuatro enfoques analíticos —descriptivo, predictivo, prescriptivo y autónomo— que representan niveles progresivos de sofisticación en el uso de la analítica, como se muestra en la Figura 1.

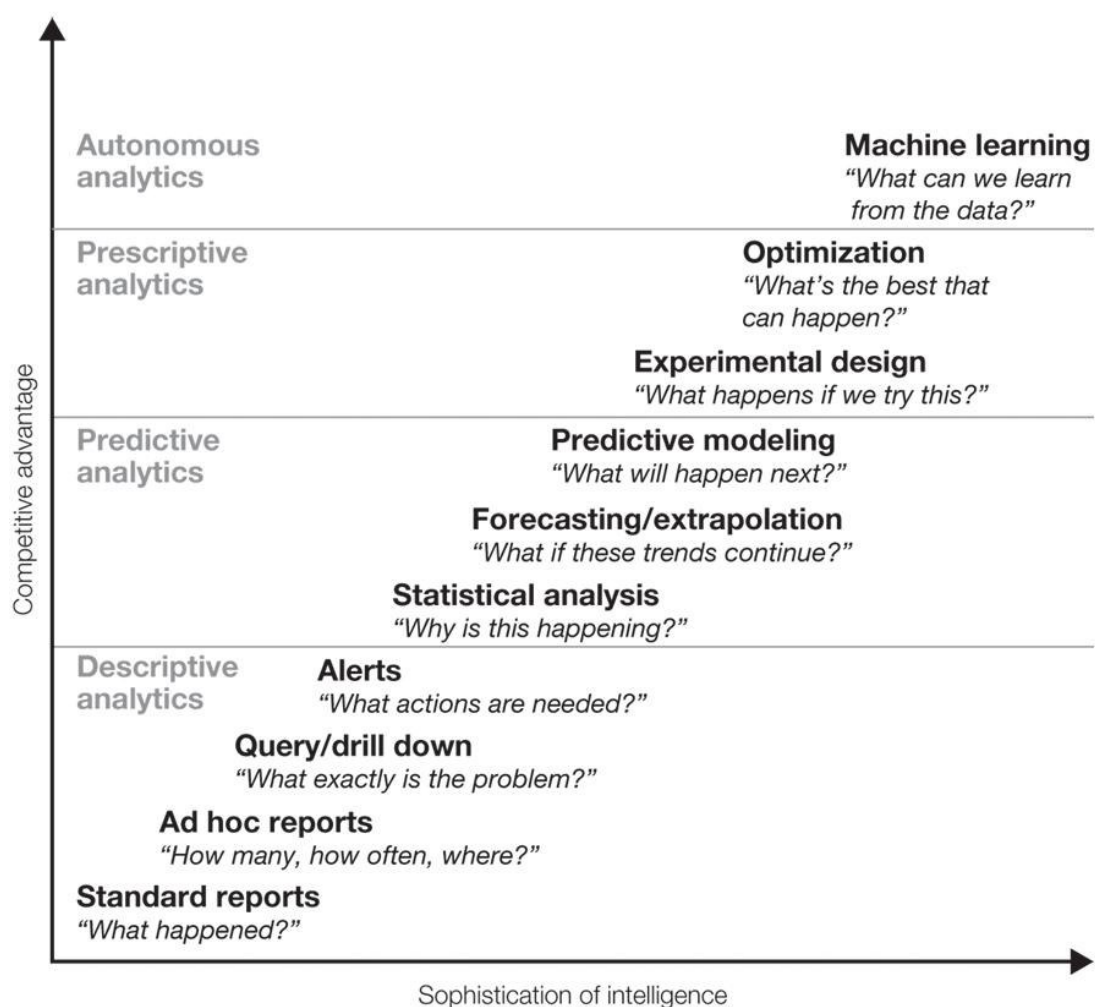


Figura 1 Niveles de sofisticación de la analítica y ventaja competitiva

Fuente. Tomado de Davenport y Harris (2007)

Sobre esta base conceptual, Ghoshal et al. (2018) presentan un marco que vincula estas capacidades analíticas con las actividades de la cadena de valor de Porter, mostrando su contribución a la creación de valor en distintas etapas de la organización.

El estudio destaca aplicaciones estratégicas tanto en actividades primarias como de apoyo, con efectos visibles en la optimización de procesos.

Bedeley et al. (2018) validan posteriormente este marco a través de estudios de caso, mostrando de forma empírica cómo las organizaciones utilizan la analítica para fortalecer su desempeño operativo, con especial énfasis en la gestión de la cadena de suministro, donde el uso estratégico de los datos se traduce en mejoras concretas en eficiencia y coordinación.

En la misma dirección, la metodología de Toma de Decisiones Basada en Datos (DDDM, por sus siglas en inglés) promueve que se emplee información de manera sistemática durante el procedimiento de tomar decisiones. En organizaciones con grandes volúmenes de datos, ayuda a tomar decisiones más acertadas al apoyarse en evidencia cuantitativa en vez de basarse solamente en la intuición (Jeyaprabha et al., 2025). Según Borges et al. (2021), este método ha resultado particularmente beneficioso en campos como el de la planificación estratégica y el marketing, pues el análisis de datos posibilita predecir situaciones y valorar alternativas con más exactitud.

2.2.6 Metodologías de Diseño, Análisis y Gestión de Datos

El modelo CRISP-DM (Cross-Industry Standard Process for Data Mining), representado en la Figura 2, es uno de los métodos más empleados para organizar proyectos de minería de datos debido a su perspectiva práctica, que se divide en seis etapas y permite su uso en diferentes ámbitos. Su validez se evidencia en investigaciones como la de Staneviciene et al. (2024), que lo emplearon para anticipar el desempeño académico en ambientes virtuales, consiguiendo detectar a tiempo a los alumnos en riesgo y posibilitando intervenciones tempranas.

Xavier et al. (2025) utilizaron el modelo CRISP-DM en un caso análogo, fusionando algoritmos de aprendizaje automático con métodos de procesamiento de lenguaje natural para automatizar la categorización de quejas en plataformas de comercio electrónico. El resultado fue una clasificación más ágil y precisa de los

comentarios, lo que contribuyó a mejorar tanto la atención al cliente como la experiencia general del usuario. El modelo alcanzó un nivel de precisión del 85%, lo que demuestra la flexibilidad y la efectividad de esta metodología en distintos tipos de proyectos analíticos.

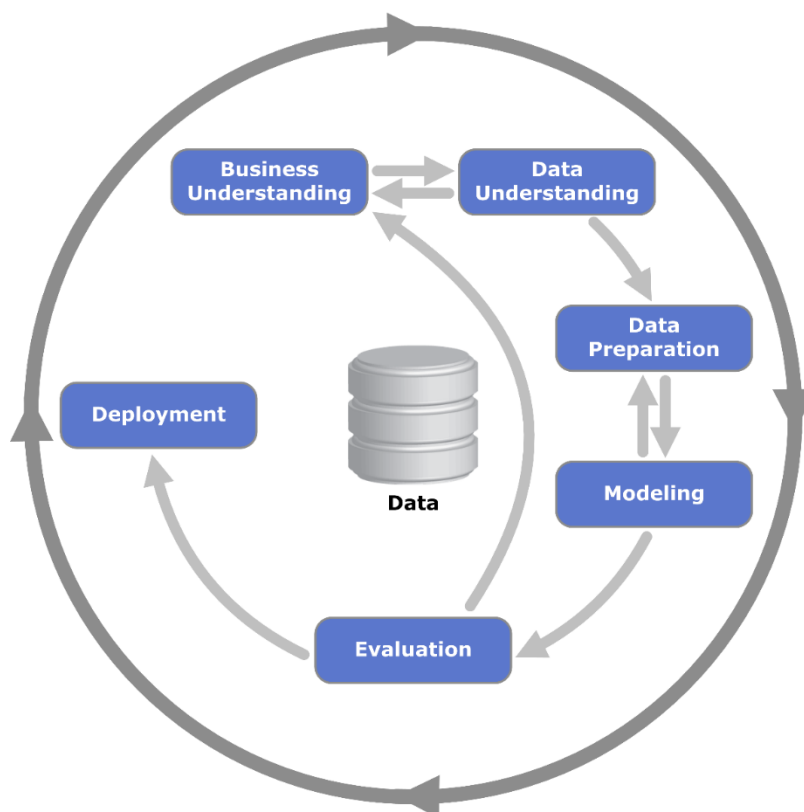


Figura 2 Fases del modelo CRISP-DM

Fuente. Tomado de Chapman et al. (2000)

En el desarrollo de soluciones de inteligencia de negocios, los enfoques propuestos por Kimball (Figura 3) e Inmon siguen siendo altamente vigentes. Kimball planteó un enfoque ascendente (bottom-up) basado en el modelado dimensional, el cual ha sido implementado con éxito en proyectos como el de bodegas peruanas orientadas a fomentar la inclusión financiera (Morales-Arevalo et al., 2025).

Por otro lado, Yessad y Labiod (2016) comparan los enfoques de Inmon, Kimball y Data Vault, y destacan que Inmon es más apropiado para arquitecturas centralizadas, estructuradas en tercera forma normal (3NF) y con una visión de consistencia a largo plazo. Kimball, en cambio, destaca por su eficiencia en el

rendimiento de consultas pensadas para usuarios finales. En contraste, Data Vault se presenta como una alternativa más flexible, ideal para entornos ágiles y con fuentes de datos que cambian constantemente.

A pesar de las diferencias entre enfoques, la metodología de Kimball sigue siendo una de las más utilizadas, principalmente porque abarca de manera integral todo el ciclo de vida de un proyecto analítico: desde la planeación hasta la entrega de valor a través del análisis de datos (Kimball et al., 2008).

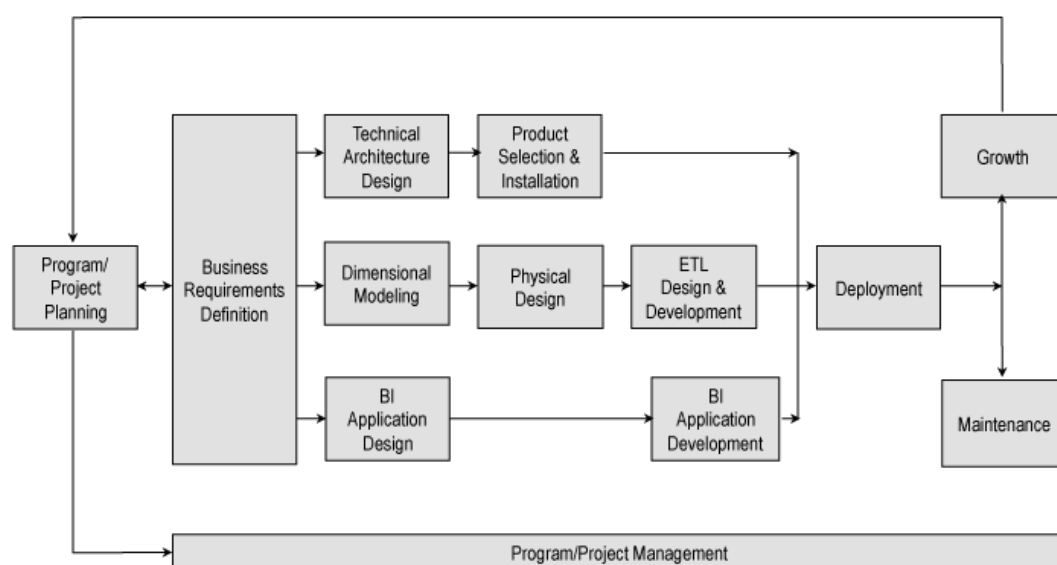


Figura 3 *Ciclo de vida de la metodología Kimball*

Fuente. Tomado de Kimball et al. (2008)

En el campo de la gobernanza de datos, el marco DAMA-DMBOK se ha convertido en una referencia fundamental para lograr una gestión integral y ordenada. Un ejemplo evidente es la arquitectura Horti-IoT, desarrollada por Amir et al. (2024) empleando esta perspectiva, lo que permitió optimizar el seguimiento en contextos agrícolas y mejorar significativamente la integración de los datos que provienen de los sensores.

Por otra parte, la metodología de DataOps tiene como objetivo automatizar cada fase del ciclo de vida de los datos, lo cual facilita que los procesos se realicen más rápidamente sin comprometer la calidad del análisis. En esta dirección, Bahaa et

al. (2023) mostraron que su implementación puede facilitar el manejo de conjuntos de datos complejos sin comprometer la exactitud de los resultados.

Con el mismo enfoque, AbouZaid et al. (2025) sugirieron un marco de trabajo fundado en Kubernetes, que hizo posible disminuir los periodos de consulta en un 12 % y optimizar la trazabilidad. Estos hallazgos muestran el efecto positivo que puede tener la implementación de modelos de DataOps en situaciones organizacionales donde el control de los datos y la agilidad son elementos fundamentales.

Para el diseño y la validación de soluciones analíticas, la metodología Design Science Research (DSR), sugerida por Peffers et al. (2007), se ha establecido como una alternativa sólida para crear y analizar artefactos científicos. Según Alnajjar et al. (2025), su estructura sistemática facilita el desarrollo de modelos de mejora, sobre todo en sectores como la construcción, donde se integran herramientas como BIM y principios de Lean Construction.

2.2.7 Calidad, Seguridad y Metadatos

ISO/IEC 25012 entrega un marco para evaluar calidad de datos. Caballero et al. (2022) lo consideran útil para medir y priorizar mejoras, y Gualo et al. (2023) reportan aplicaciones industriales.

Para metadatos, ISO/IEC 11179 favorece interoperabilidad semántica; Hegselmann et al. (2021) validan su eficacia en repositorios médicos y Hu et al. (2024) en integraciones interinstitucionales. ISO/IEC 27001 ofrece un marco de gestión de riesgos y cumplimiento en seguridad de la información (Malatji, 2023). En Big Data, ayuda a equilibrar privacidad y aprovechamiento analítico (Utomo et al., 2024).

ISO 8000 es clave para datos maestros y golden records, alineada con ISO/IEC 25012, lo que reduce redundancias e impulsa la interoperabilidad en estructuras complejas (Gualo et al., 2023; Miller et al., 2025).

2.2.8 Visualización de Datos

La visualización mediante dashboards y KPIs es esencial para decisiones en tiempo real. Aradiel Castañeda et al. (2024) muestran que tableros sumados a RPA incrementan la eficiencia operativa.

Nunes et al. (2024) evidencian que estandarizar KPIs mejora la base de las decisiones estratégicas.

El storytelling visual gana terreno para comunicar hallazgos de forma clara y persuasiva. Komóczy y Kósa (2025) reportan su utilidad en salud mental y Song et al. (2025) exploran su empleo en cine.

En analítica visual, Han et al. (2025) desarrollan VNE para detectar zonas urbanas vulnerables a desastres y Yan et al. (2025) estudian IA generativa para mejorar comprensión estudiantil de visualizaciones. En conjunto, se mantiene la necesidad de un modelo integral que articule estas herramientas bajo una estructura flexible y adaptativa.

2.3 Bases Teóricas

Kahneman y Tversky (2011) describen la decisión como evaluación de opciones bajo incertidumbre, con resultados esperados de ganancia o pérdida.

Amann y Gómez (2014) proponen un enfoque holístico que observe el sistema completo y sus interacciones para identificar patrones y causas estructurales.

Almeida y Gama (2017) amplían esta propuesta al destacar la convergencia de disciplinas como estadística, informática y ciencias sociales para abordar problemas complejos.

La toma de decisiones no es más que incluir el proceso mediante el cual un individuo o toda una organización pretenden llegar a una conclusión sobre qué

acciones adicionales emprender en el futuro con un conjunto de objetivos y límites en los recursos accesibles. La decisión es literalmente una elección, es decir, cuando un individuo tiene la intención de llegar a una conclusión con respecto a circunstancias o situaciones particulares. Denota un curso de comportamiento, es decir, lo que uno debería hacer o no hacer. Por otro lado, la toma de decisiones como proceso se refiere a la selección de una acción particular de múltiples cursos de acción. Por lo tanto, es una función de toma de decisiones y la elección que eventualmente determina la acción o inacción de uno. Los procesos involucrados durante la toma de decisiones generalmente están influenciados por varios factores, es decir, factores internos o externos (Manoharan et al., 2023).

Según ISO 31010 (2019), la Incertidumbre en las decisiones es la que tiene especial relevancia para las estrategias de gestión de riesgos y que identifica la incertidumbre asociada con los sistemas de valores, el juicio profesional, los valores de la empresa y las normas sociales.

Según Benítez (2018), Business intelligence o también llamado La Inteligencia de Negocios, hace referencia al manejo optimizado de los que almacena, recopila y analiza una organización, siendo capaz de transformarlos en decisiones estratégicas que permitan el diseño de acciones orientadas a alcanzar el éxito empresarial.

Business Analytics es la ciencia de plantear y responder preguntas de datos relacionadas con los negocios. El análisis empresarial se ha expandido rápidamente en los últimos años para incluir herramientas extraídas de estadísticas, gestión de datos, visualización de datos y aprendizaje automático (Pochiraju & Seshadri, 2019).

Big data puede tener los siguientes activos como alta velocidad, gran volumen y gran variedad que implican un procesamiento que ayuda en la toma de decisiones y la optimización de procesos. Big data es una expresión que significa gran volumen de datos que pueden ser estructurados o no estructurados y cuyo procesamiento es muy difícil utilizando una base de datos tradicional (Kulkarni et al., 2019).

El término Inteligencia Artificial (IA) fue acuñado por John McCarthy en 1956.

Los científicos e investigadores han propuesto numerosas definiciones para la IA, entre ellas, el estudio orientado a desarrollar sistemas computacionales capaces de realizar tareas que, hasta ahora, requieren de habilidades propias del ser humano; la IA constituye un campo de la informática enfocado en el diseño de sistemas inteligentes, es decir, aquellos que manifiestan comportamientos comúnmente asociados a la inteligencia humana; asimismo, se la considera una rama de la informática dedicada a la representación del conocimiento mediante símbolos, así como al uso de reglas empíricas y métodos heurísticos para el procesamiento de la información (Gupta & Mangla, 2020).

La efectividad se refiere a la capacidad de alcanzar los resultados deseados con el menor esfuerzo o costo posible. En otras palabras, evalúa en qué medida las acciones emprendidas logran cumplir los objetivos establecidos y generan un valor tangible, ya sea para una organización o para una persona (Drucker, 1967).

En el campo de la investigación en ciencias del diseño (DSR), se entiende por pertinencia que el artefacto creado esté en consonancia con las necesidades auténticas del contexto, lo cual se organiza a través del ciclo de relevancia que Hevner (2007) propuso. Desde este punto de vista, la relevancia se comprende como la habilidad de una solución para abordar problemas concretos y dar respuesta a las exigencias particulares del entorno en el que se aplica.

Cuando un artefacto es capaz de convertir datos en conocimiento que se puede aplicar, ayudando a disminuir la incertidumbre en los niveles gerenciales y respaldando las decisiones estratégicas, se verifica su utilidad (Gilliot et al., 2024).

La viabilidad está vinculada a factores importantes, como la infraestructura tecnológica existente, la calidad de los datos y las barreras culturales que podrían interferir con su utilización e implementación. En relación a esto, Pratt et al. (2020) afirman que el grado de madurez tecnológica es determinante para la adopción exitosa.

Finalmente, la eficacia del diseño se refleja en qué tan bien los resultados obtenidos se alinean con los objetivos propuestos. Siguiendo la metodología de Peffers

et al. (2007), la eficacia no se asume como un hecho teórico, sino que debe comprobarse mediante una evaluación empírica, contrastando lo que se esperaba lograr con los efectos reales del modelo en su contexto de aplicación. Es este cumplimiento de objetivos el que termina por legitimar la solidez del modelo holístico frente a las exigencias del entorno organizacional.

2.4 Glosario de términos.

ADA – Advanced Data Analytics

ADA-CMM – Advanced Data Analytics Capability Maturity Model

ACDGE – Advanced Capability-Driven Governance and Evaluation Model

AML – Analytic Maturity Level

APMM – Analytic Process Maturity Model

BA – Business Analytics

BI – Business Intelligence

BIA – Business Intelligence and Analytics

CMM – Capability Maturity Model

CRISP-DM – Cross-Industry Standard Process for Data Mining

DQ – Data Quality

DSR / DSRM – Design Science Research / Design Science Research Methodology

IES – Instituciones de Educación Superior

IT / TI – Information Technology / Tecnologías de la Información

KDD – Knowledge Discovery in Databases

KPI – Key Performance Indicator

PLS-SEM – Partial Least Squares Structural Equation Modeling

PYMES / SMEs – Pequeñas y Medianas Empresas / Small and Medium-sized Enterprises

TD – Toma de Decisiones

SIS – Seguro Integral de Salud (por el caso aplicado con datos abiertos)

CAPÍTULO 3: METODOLOGÍA

La metodología de investigación permite articular conocimientos provenientes de disciplinas como la estadística, la analítica avanzada, la inteligencia artificial y la gestión organizacional. En entornos caracterizados por elevada complejidad decisional, los enfoques unidimensionales tienden a revelar sus límites; en su lugar, se vuelve necesaria una articulación interdisciplinaria capaz de estructurar soluciones integrales sustentadas en datos (Provost & Fawcett, 2013).

La presente investigación se inscribe en el paradigma de Design Science Research (DSR), cuyo núcleo epistemológico reside en el diseño, la construcción y la evaluación sistemática de artefactos destinados a afrontar problemas organizacionales complejos. Bajo este enfoque, el rigor científico no se reduce a la contrastación estadística de hipótesis, sino que se expresa en la evaluación estructurada de la utilidad, la pertinencia contextual, la viabilidad técnica y la eficacia operativa del artefacto propuesto (Hevner et al., 2004; Peffers et al., 2007; Hevner & Chatterjee, 2010).

En coherencia con esta lógica, la investigación formula proposiciones de diseño entendidas como enunciados evaluativos que orientan la configuración, especificación y validación del Modelo Holístico de Analítica Avanzada. Tales proposiciones no buscan establecer relaciones causales universalmente generalizables; su propósito consiste, más bien, en demostrar que el artefacto desarrollado satisface su finalidad declarada: fortalecer la toma de decisiones organizacionales mediante una arquitectura analítica integrada.

Las dimensiones adoptadas —pertinencia, viabilidad, utilidad y eficacia operativa— operan como criterios analíticos para valorar el desempeño del modelo, tanto a través del juicio estructurado de expertos como mediante su implementación en un caso aplicado. Sobre esta base, el capítulo detalla el diseño metodológico implementado, las técnicas de recolección y análisis empleadas y los procedimientos de validación desplegados para examinar el artefacto propuesto.

3.1 Tipo y Diseño de Investigación

3.1.1 *Tipo de Investigación*

El presente estudio es de naturaleza aplicada, ya que aborda un problema práctico: optimizar la toma de decisiones organizativas mediante el diseño y la validación de un modelo holístico de análisis avanzado. Según Creswell y Creswell (2018), la investigación aplicada se orienta a trasladar el conocimiento teórico hacia contextos empíricos concretos, procurando generar soluciones con impacto verificable en entornos específicos más que limitarse a la producción conceptual.

Desde una perspectiva metodológica, el estudio se inscribe en la investigación en ciencias del diseño (Design Science Research, DSR), dado que este enfoque resulta particularmente idóneo para el diseño, la construcción y la evaluación de artefactos —modelos, métodos o sistemas— concebidos para abordar problemas organizativos complejos. La DSR no disocia rigor y aplicabilidad; por el contrario, los articula, en la medida en que exige fundamentación teórica sólida a la vez que demanda validación práctica del conocimiento producido (Hevner et al., 2004; Peffers et al., 2007).

En coherencia con este encuadre, se adopta el modelo clásico de toma de decisiones propuesto por Simon (1960), estructurado en las fases de inteligencia, diseño y elección, como marco conceptual para interpretar el proceso decisional organizativo. Aunque no constituye la metodología del estudio, dicho modelo opera como sustrato teórico que permite alinear el artefacto analítico desarrollado con las dinámicas reales de decisión.

El artefacto se configura a partir de una revisión sistemática de la literatura y un análisis de contenido estructurado, integrando marcos teóricos consolidados, modelos analíticos preexistentes y estándares internacionales vinculados a la gobernanza y la calidad de los datos. Su validación se realiza mediante juicio experto y estudio de caso aplicado, conforme a los principios evaluativos propios de la investigación en ciencias del diseño (Hevner y Chatterjee, 2010; Van Aken, 2004).

3.1.2 Enfoque metodológico: Design Science Research (DSR)

La DSR es el fundamento metodológico de esta investigación, ya que brinda un marco lógico y bien estructurado para orientar la creación, el desarrollo y la validación del modelo de análisis avanzado, presentado como artefacto científico. Como se ilustra en la Figura 4, este proceso iterativo tiene seis etapas: identificación del problema, definición de objetivos, diseño y desarrollo, demostración, evaluación y comunicación. Esta secuencia es discutida por Peffers et al. (2007) y desarrollada por Hevner & Chatterjee (2010).

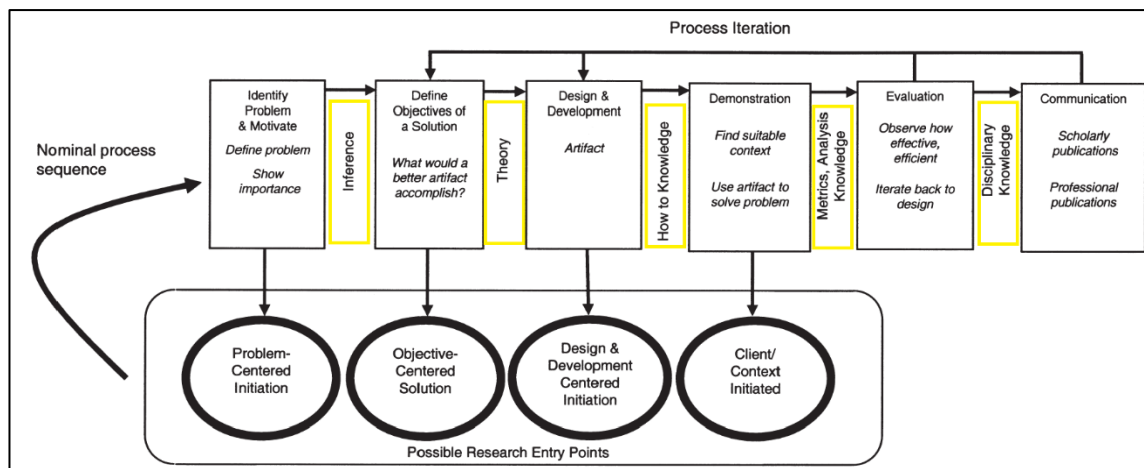


Figura 4 DSRM Process Model

Fuente. Tomado de Hevner & Chatterjee (2010)

3.1.3 Diseño de la investigación basado en DSR

De acuerdo con Hevner y Chatterjee (2010), el modelo DSRM se estructura en seis fases que guían el diseño y validación de artefactos científicos en sistemas de información:

3.1.3.1 Identificación del problema y motivación

Destaca que esta fase requiere definir con claridad el problema de investigación y justificar el valor de su solución; el análisis conceptual del problema permite comprender su complejidad, mientras que la justificación motiva tanto al investigador

como al público; el autor destaca la necesidad de comprender el estado actual del problema y la relevancia de su solución (pp. 28-29).

3.1.3.2 Definición de los objetivos de la solución

Los objetivos deben originarse racionalmente del problema identificado, teniendo en cuenta lo técnicamente factible; estos pueden ser cuantitativos (por ejemplo, una mejora métrica) o cualitativos (por ejemplo, abordar problemas sin resolver); el autor destaca la importancia de alinear estos objetivos con los requisitos del artefacto, incluso cuando las soluciones son parciales o incrementales (p. 29).

3.1.3.3 Diseño y desarrollo

Caracterizar esta fase como la creación de artefactos (constructos, modelos, métodos o instanciaciones) que incorporan las contribuciones de la investigación; el autor destaca que el proceso requiere definir tanto la funcionalidad como la arquitectura del artefacto, basándose en las teorías existentes para fundamentar la solución (p. 29).

3.1.3.4 Demostración

Esta fase consiste en demostrar la aplicación del artefacto mediante experimentos, simulaciones, estudios de casos o pruebas concretas; advierte que la demostración requiere una comprensión profunda de cómo aplicar el artefacto para resolver el problema definido (p. 30).

3.1.3.5 Evaluación

El autor propone comparar los resultados observados con los objetivos establecidos, utilizando métricas cuantitativas (por ejemplo, el rendimiento) o cualitativas (por ejemplo, los comentarios de los usuarios). Destaca que esta fase puede

dar lugar a iteraciones para perfeccionar el artefacto o, si tiene éxito, para comunicar los resultados (p. 30).

3.1.3.6 Comunicación

El autor propone difundir el problema, el artefacto y su eficacia entre el público académico y el profesional. Sugiere que las publicaciones pueden estructurarse siguiendo el flujo del DSRM, al igual que los estudios empíricos siguen un formato estándar (p. 30).

3.1.4 Metodología para el Proceso de Toma de Decisiones

De acuerdo con Simon (1960), el proceso decisional se articula en tres fases fundamentales: inteligencia, diseño y elección. Tal como se representa en la figura 5, este esquema no debe interpretarse como una secuencia lineal cerrada, sino como una dinámica iterativa que ofrece una matriz estructural para abordar problemas organizativos de elevada complejidad.

Evidencia empírica reciente ha reafirmado la vigencia de este marco. Liberatore y Wagner (2022), al analizar la dinámica interna de los procesos decisionales, muestran que la asignación de esfuerzo entre las fases dista de ser homogénea: la fase de diseño —orientada a la generación y estructuración de alternativas— demanda una inversión temporal y cognitiva significativamente mayor que las fases de inteligencia y elección.

Desde una lectura crítica, la persistencia empírica del modelo de Simon en contextos diversos sugiere una consistencia teórica difícil de ignorar. La fase de diseño aparece como componente transversal, estable a través de distintos tipos de tarea, configuraciones tecnológicas y sectores organizativos; su centralidad refuerza la pertinencia del modelo como representación del comportamiento decisional gerencial y subraya el valor estratégico de fortalecer los mecanismos de formulación y evaluación de alternativas.

La recurrencia de estas fases en múltiples entornos organizativos apunta a una lógica estructural subyacente que trasciende contingencias operativas específicas. En este sentido, alinear arquitecturas analíticas avanzadas con dicha estructura decisional no constituye una mera elección instrumental, sino una apuesta teóricamente fundamentada para mejorar —y potencialmente optimizar— los procesos de toma de decisiones organizativas.

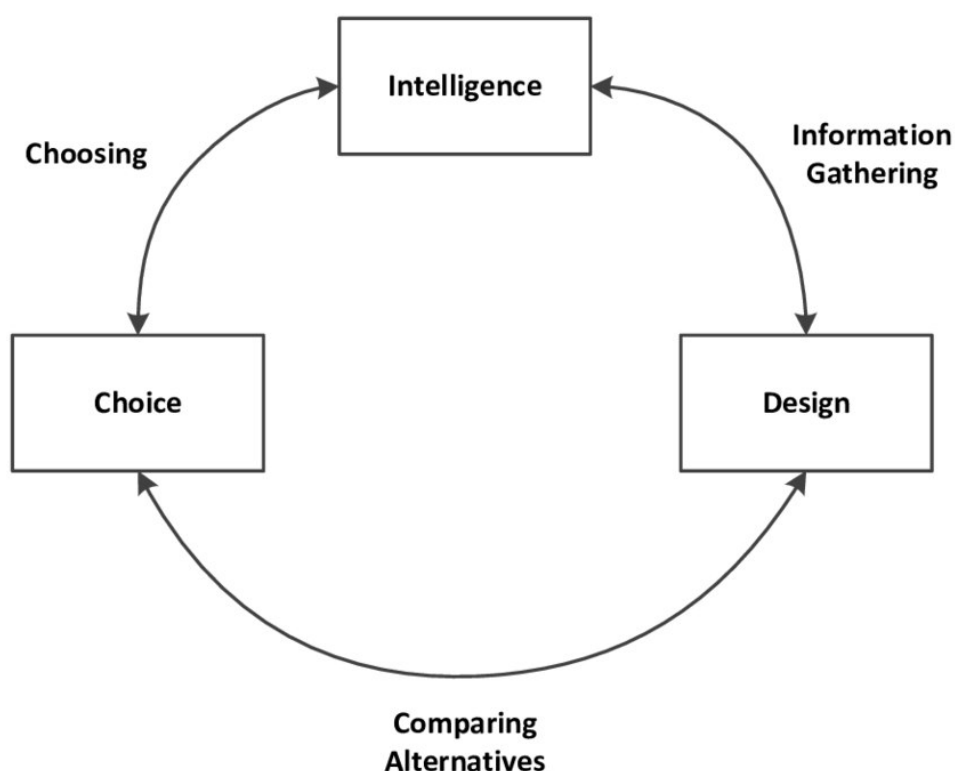


Figura 5 *Modelo de proceso de toma de decisiones en tres fases*

Fuente. Tomado de Simon (1960)

De acuerdo con Simon (1960), el proceso de toma de decisiones se estructura en tres fases: inteligencia, diseño y elección:

3.1.4.1 Inteligencia

Define esta fase inicial como el proceso de examinar minuciosamente el entorno para identificar situaciones que requieren una decisión. El autor utiliza el

término «inteligencia» en su sentido militar, haciendo hincapié en la recopilación sistemática de información crítica que sirve de base para la acción (p. 2).

3.1.4.2 Diseño

Explica cómo se generan y analizan posibles cursos de acción. El diseño implica creatividad (invención de alternativas) y rigor (evaluación de su viabilidad), y constituye un puente entre la identificación del problema y la selección final (p. 2).

3.1.4.3 Elección

Describe esta fase como la selección de una alternativa específica entre las alternativas disponibles. Destaca que la elección no es meramente intuitiva, sino que debe basarse en los criterios establecidos en las fases previas de inteligencia y diseño (p. 2).

3.2 Propositiones

3.2.1 Proposición general

El desarrollo de un modelo holístico de analítica avanzada optimizará la toma de decisiones organizacionales.

3.2.2 Propositiones específicas

- PR1: La estructuración de las fases del Modelo Holístico de Analítica Avanzada mediante la integración sistemática de marcos teóricos y el análisis sistemático de la literatura, permitirá establecer la arquitectura del artefacto analítico.
- PR2: El diseño del Modelo Holístico de Analítica Avanzada mediante la descomposición funcional de su arquitectura, especificando componentes y flujos de integración, permitirá su articulación con el proceso de toma de decisiones

organizacionales.

- PR3: La validación del modelo holístico, mediante el juicio de expertos, evidenciará altos niveles de pertinencia, utilidad y viabilidad, y, mediante un caso aplicado, demostrará su eficacia para optimizar la toma de decisiones organizacionales.

3.3 Dimensiones del modelo y criterios de evaluación

Seguidamente, se describen las dimensiones conceptuales del modelo holístico de analítica avanzada, las cuales se utilizan como componentes estructurales del artefacto y como criterios para su evaluación, de acuerdo con el enfoque de Design Science Research.

3.3.1 Identificación de dimensiones del modelo

3.3.1.1 Dimensión de toma de decisiones organizacionales

Corresponde al proceso de toma de decisiones en el contexto organizacional, considerado como el dominio de aplicación del artefacto analítico. Esta dimensión permite evaluar en qué medida el modelo propuesto se articula con las etapas del proceso decisional y contribuye a optimizarlo.

3.3.1.2 Dimensión de analítica avanzada

Representa el conjunto de capacidades analíticas incorporadas en el modelo holístico, incluyendo los niveles descriptivo, predictivo, prescriptivo y autónomo. Esta dimensión constituye el núcleo del artefacto analítico y define los mecanismos mediante los cuales se generan insights para apoyar la toma de decisiones organizacionales.

3.3.2 Operacionalización de criterios de evaluación

La operacionalización, presentada en la Tabla 2, es clave para que la analítica avanzada cumpla su propósito y fortalezca la decisión. Definiciones vagas reducen impacto y utilidad. Sin un esquema claro, muchos proyectos quedan en exploración y no llegan a soluciones productivas. Gartner (2022) advierte que la ausencia de una operacionalización sostenible puede retrasar la implementación de iniciativas de datos por largos periodos, lo que subraya la necesidad de procesos sistemáticos que traduzcan modelos en soluciones implementables.

Tabla 2

Matriz de Operacionalización de criterios de evaluación del modelo

Dimensiones del Modelo	Definición Conceptual	Definición Operacional	Criterio de Evaluación	Indicadores	Instrumento / técnica de evaluación
Toma de decisiones organizacionales	Proceso mediante el cual las organizaciones seleccionan un curso de acción entre varias alternativas para lograr sus objetivos.	Nivel de efectividad, rapidez y precisión en la selección de opciones y ejecución de acciones para alcanzar objetivos empresariales.	Pertinencia	Coherencia del modelo con el proceso de toma de decisiones organizacionales	Likert (1–5) – Juicio de expertos
			Utilidad	Contribución del modelo a la mejora del proceso decisional.	Likert (1–5) – Juicio de expertos
			Viabilidad	Factibilidad de implementación del modelo en contextos	Likert (1–5) – Juicio de expertos

Analítica avanzada	Uso de técnicas y herramientas avanzadas de análisis de datos para apoyar y mejorar el proceso de toma de decisiones organizacionales.	Grado en que las organizaciones implementan y utilizan técnicas avanzadas de análisis de datos para la toma de decisiones.	Eficacia	organizacionales reales Mejora en resultados clave, reducción de errores, cumplimiento de plazos	Comparación antes/después – Caso aplicado
			Analítica Descriptiva	Uso de reportes, frecuencia de análisis, comprensión de información	Likert (1–5) – Juicio de expertos
			Analítica Predictiva	Uso de modelos predictivos Precisión de los modelos predictivos	Métricas analíticas (%) – Caso aplicado
			Analítica Prescriptiva	Aplicación de recomendaciones analíticas en decisiones reales	Métricas analíticas (%) – Caso aplicado
				Aplicación de recomendaciones analíticas en decisiones reales	Evidencia de implementación – Caso aplicado

		Grado de automatización en la toma de decisiones basada en análisis	Likert (1–5) – Juicio de expertos
	Analítica Autónoma	Número de decisiones tomadas automáticamente	Escala de frecuencia (número de decisiones)

3.4 Matriz de Consistencia

La transición del plano abstracto hacia la mensurabilidad empírica se cristaliza en la operacionalización detallada en la Tabla 3, un componente cuya relevancia trasciende la organización formal de la investigación para erigirse en el eje articulador de la analítica avanzada. Al delimitar con rigor las variables, dimensiones e indicadores, se dota al modelo de la capacidad de transformar constructos teóricos en resultados tangibles y auditables; sin esta precisión, el análisis de datos corre el riesgo de diluirse en la ambigüedad conceptual. La carencia de una ruta crítica que defina cómo se cuantifican los fenómenos bajo estudio es, con frecuencia, la causa raíz del estancamiento de las iniciativas basadas en datos, las cuales suelen perder coherencia operativa al no contar con un anclaje métrico sólido.

Bajo esta perspectiva, la advertencia de Gartner (2022) cobra una vigencia alarmante en el contexto organizacional: la ausencia de un marco de operacionalización sostenible no solo compromete la integridad del proyecto, sino que puede derivar en ineficiencias críticas, extendiendo los plazos de implementación de soluciones analíticas hasta por un periodo de dos años. Este escenario de obsolescencia prematura y retraso sistémico refuerza la necesidad imperativa de adoptar procesos estructurados. Solo mediante una transición metódica y bien documentada entre el análisis de datos y su despliegue efectivo es posible garantizar que el modelo holístico cumpla su propósito de optimizar la toma de decisiones, convirtiendo la arquitectura informacional en una verdadera ventaja competitiva para la institución:

Tabla 3
Matriz de Consistencia

PROBLEMAS	OBJETIVOS	PROPOSICIONES	DIMENSIONES
Problema general	Objetivo general	Proposición general	

¿Cómo el desarrollo de un modelo holístico de analítica avanzada optimiza la toma de decisiones organizacionales?	Desarrollar un modelo holístico de analítica avanzada para optimizar la toma de decisiones organizacionales.	El desarrollo de un modelo holístico de analítica avanzada optimizará la toma de decisiones organizacionales.
---	--	---

Problemas específicos	Objetivos específicos	Proposiciones específicas	Dimensión de Toma de decisiones organizacionales.
¿Cómo la estructuración de las fases del Modelo Holístico de Analítica Avanzada, mediante la integración sistemática de marcos teóricos y el análisis sistemático de la literatura, establece la arquitectura del artefacto analítico?	Estructurar las fases del Modelo Holístico de Analítica Avanzada mediante la integración sistemática de marcos teóricos y el análisis sistemático de la literatura, para establecer la arquitectura del artefacto analítico.	La estructuración de las fases del Modelo Holístico de Analítica Avanzada mediante la integración sistemática de marcos teóricos y el análisis sistemático de la literatura permitirá establecer la arquitectura del artefacto analítico.	Dimensión de Analítica avanzada.
¿Cómo el diseño del Modelo Holístico de Analítica Avanzada mediante la descomposición funcional de su arquitectura, especificando componentes y flujos de integración, se articula con el proceso de toma de decisiones organizacionales?	Diseñar el Modelo Holístico de Analítica Avanzada mediante la descomposición funcional de su arquitectura, especificando componentes y flujos de integración para su articulación con el proceso de toma de decisiones organizacionales.	El diseño del Modelo Holístico de Analítica Avanzada mediante la descomposición funcional de su arquitectura, especificando componentes y flujos de integración, permitirá su articulación con el proceso de toma de decisiones organizacionales.	

¿Cómo el diseño del Modelo Holístico de Analítica Avanzada mediante la descomposición funcional de su arquitectura, especificando componentes y flujos de integración, se articula con el proceso de toma de decisiones organizacionales?	Validar el modelo holístico de analítica avanzada mediante el juicio de expertos para evidenciar altos niveles de pertinencia, utilidad y viabilidad, y mediante un caso aplicado para demostrar su eficacia en la optimización de la toma de decisiones organizacionales.	La validación del modelo holístico, mediante el juicio de expertos, evidenciará altos niveles de pertinencia, utilidad y viabilidad, y, mediante un caso aplicado, demostrará su eficacia para optimizar la toma de decisiones organizacionales.
---	--	--

3.5 Unidad de análisis

La literatura científica especializada, que incluye marcos de referencia, modelos conceptuales, artículos académicos, estándares y regulaciones internacionales relacionados con la toma de decisiones a nivel organizacional y la analítica avanzada, forma parte del objeto de estudio del presente trabajo.

Acorde con la perspectiva de Design Science Research, el análisis no se enfoca directamente en las entidades o en los individuos, sino que se concentra en la síntesis, interpretación e integración del conocimiento ya existente. Este último es el fundamento a partir del cual se organiza el modelo holístico sugerido como artefacto científico (Peffer et al., 2007; Hevner et al., 2004).

3.6 Población y muestra

3.6.1 Población de estudio

La población de referencia está conformada por publicaciones científicas, modelos conceptuales, marcos metodológicos y estándares internacionales relacionados con la analítica avanzada y la toma de decisiones organizacionales.

No se busca realizar inferencias estadísticas poblacionales, sino fundamentar el diseño del modelo a partir de la evidencia teórica y metodológica acumulada en la literatura científica especializada.

3.6.2 Tamaño de la muestra

Dado que la investigación no persigue objetivos inferenciales, no se define un tamaño de muestra estadístico tradicional. La información empírica recolectada cumple un rol diagnóstico y contextual, orientado a comprender el estado general del uso de la analítica avanzada, sin pretensión de generalización.

En la fase de validación, el tamaño de la muestra se determina por criterios de

suficiencia teórica y saturación, particularmente en el juicio de expertos, conforme a las recomendaciones metodológicas de la Design Science Research (Hevner & Chatterjee, 2010)

3.6.3 Selección de muestra

La selección de los participantes se realizó mediante muestreo intencional, priorizando perfiles con experiencia en analítica, toma de decisiones, gestión de datos y sistemas de información. En el caso del juicio de expertos, se consideraron criterios como trayectoria académica, experiencia profesional y conocimiento del dominio, siguiendo prácticas habituales en estudios de validación de modelos (Okoli & Pawlowski, 2004).

3.7 Técnicas de recolección de Datos

En coherencia con el enfoque cualitativo–propositivo adoptado bajo el paradigma de Design Science Research, la investigación emplea técnicas cualitativas y cuantitativas con propósitos claramente diferenciados:

3.7.1 Revisión sistemática de la literatura

Empleada como método cualitativo principal para la creación del modelo holístico, facilita el reconocimiento de los fundamentos teóricos, las lagunas presentes y las buenas prácticas en la analítica avanzada y en la toma de decisiones organizacionales.

3.7.2 Juicio de expertos

Se recurrió al juicio de expertos con el propósito de determinar la pertinencia, utilidad y viabilidad del artefacto propuesto en contextos organizacionales reales. Esta modalidad evaluativa aporta una perspectiva cualitativa que complementa los resultados analíticos y respalda la validación práctica en escenarios decisionales donde la información disponible puede presentarse incompleta, ambigua o fragmentada; en el marco de la investigación en ciencias del diseño (DSR), tal instancia resulta decisiva para constatar que el artefacto cumple con su finalidad proyectada y exhibe aplicabilidad

efectiva.

Cuando se implementa de manera estructurada, el juicio experto configura un mecanismo evaluativo metodológicamente consistente, tal como lo evidencian antecedentes empíricos. Buitrago Pulido et al. (2023) destacan la importancia de incorporar especialistas provenientes de distintas áreas en la validación del modelo ACDGE, con el fin de examinar su validez de contenido y coherencia interna; sus resultados muestran un elevado nivel de concordancia entre evaluadores y sugieren que la retroalimentación experta no solo confirma la solidez estructural del modelo, sino que también permite identificar ajustes potenciales que fortalecen su configuración final.

Por su parte, Stander et al. (2023) subrayan la importancia de las opiniones de expertos en decisiones técnicas complicadas, como es el caso de elegir tecnologías para ambientes industriales. Aseguran que, aunque los cuestionarios cerrados poseen ciertas limitaciones, al fusionarse con valoraciones cualitativas de expertos se logran resultados más exactos y con menos prejuicios. Esto es especialmente beneficioso en proyectos de análisis avanzado, en los que participan elementos organizacionales, técnicos y contextuales que exigen una interpretación más allá de lo puramente numérico.

Se realizó la evaluación del modelo holístico sugerido a través de un cuestionario estructurado orientado a expertos, basándose en lo anterior. Este instrumento se elaboró considerando dimensiones de evaluación que han sido reconocidas en la literatura científica, lo cual asegura una congruencia conceptual y posibilita la comparación entre las respuestas. El cuestionario incorpora componentes de los siguientes marcos de referencia:

- **Modelo de Aceptación de la Tecnología (TAM)**, que permite evaluar la **utilidad percibida** y la **facilidad de uso**, factores clave para la adopción de soluciones analíticas en contextos organizacionales (Mikša & Sikirica, n.d.).
- **System Usability Scale (SUS)**, utilizada para medir de forma fiable la **comprensión**, la **facilidad de interacción** y la **usabilidad general del modelo**, incluso en etapas tempranas del diseño conceptual (Harrati et al., 2016).

- **Método Delphi y juicio de expertos**, que proporcionan un marco sistemático para alcanzar consenso entre especialistas y son especialmente eficaces en la evaluación de modelos conceptuales y decisiones organizacionales complejas (Guillemot et al., 2024).

La combinación de estos enfoques permitió construir un instrumento de validación robusto, orientado a evaluar de manera integral la **pertinencia, utilidad y viabilidad** del modelo holístico de analítica avanzada, garantizando así el rigor metodológico del proceso de evaluación conforme a los principios de la Design Science Research.

La Tabla 4 muestra el conjunto de preguntas formuladas como instrumento de validación, junto con los criterios evaluados en cada una de ellas.

Tabla 4

Preguntas del instrumento de validación y criterios evaluados

Nº	Pregunta	Criterio evaluado
1	¿El modelo propuesto refleja adecuadamente las fases clave de un proceso de toma de decisiones basado en datos?	Pertinencia
2	¿Las fases del modelo están claramente definidas y son comprensibles para su implementación en una organización?	Viabilidad
3	¿El modelo se adapta a diferentes contextos organizativos sin necesidad de cambios estructurales?	Viabilidad
4	¿El modelo integra adecuadamente marcos y metodologías reconocidos (por ejemplo, ISO 25012, DSR, modelo de proceso TD)?	Pertinencia
5	¿Las relaciones entre las fases del modelo son coherentes y lógicamente secuenciales?	Pertinencia
6	¿El modelo puede ser replicado por otros investigadores o profesionales con una ambigüedad mínima?	Viabilidad
7	¿Las capacidades analíticas descritas (descriptivas, predictivas, prescriptivas, autónomas) están correctamente clasificadas?	Pertinencia
8	¿La inclusión de la fase de visualización (KPI, narración de historias, paneles de control) contribuye a la comprensión y el uso de los resultados?	Utilidad

9	¿El modelo añade valor a la toma de decisiones organizativas desde un enfoque holístico e integrador?	Utilidad
10	¿Creo que el modelo propuesto es útil y aplicable para mejorar los procesos de toma de decisiones basados en análisis avanzados?	Utilidad

Fuente. Adaptado de TAM, SUS y Delphi.

3.7.3 Caso aplicado

Utilizado como técnica de demostración cuantitativa descriptiva del modelo, mediante la aplicación de algoritmos de analítica avanzada implementados en Python sobre un conjunto de datos reales. Este procedimiento tiene un carácter ilustrativo y demostrativo, orientado a evidenciar la aplicabilidad operativa del modelo, sin pretensión de validación estadística inferencial ni de contrastación de hipótesis.

3.8 Procesamiento y análisis de datos

El procesamiento y análisis de la información se realiza desde una perspectiva cualitativa–analítica, en coherencia con el enfoque de Design Science Research adoptado, diferenciando tres niveles de análisis, cada uno alineado con los objetivos específicos de la investigación:

3.8.1 Análisis de contenido de la literatura científica

Se realiza un análisis sistemático y comparativo de la literatura científica, modelos, marcos de referencia y estándares internacionales, con el propósito de identificar patrones, vacíos conceptuales y elementos clave que fundamentan la estructuración del modelo holístico de analítica avanzada. Este análisis permite integrar conceptualmente los componentes del modelo y asegurar su alineación con el estado del arte.

3.8.2 Análisis de los resultados del juicio de expertos

Los resultados del juicio de expertos se analizan mediante técnicas de síntesis

cualitativa y de análisis descriptivo, orientadas a evaluar la pertinencia, la utilidad y la viabilidad del modelo propuesto. Este análisis se realiza conforme a los principios de evaluación establecidos en la Design Science Research, lo que permite identificar niveles de consenso, observaciones críticas y oportunidades de mejora del artefacto diseñado.

3.8.3 Análisis del caso aplicado mediante datos abiertos

En la fase de demostración, se analizan los resultados generados mediante algoritmos de analítica avanzada implementados en Python, aplicados a un conjunto de datos abiertos. Este ejercicio tiene un propósito práctico y demostrativo, enfocado en mostrar la aplicabilidad del modelo holístico en un contexto real de toma de decisiones, sin buscar validación estadística inferencial ni generalizar los resultados.

CAPÍTULO 4: DESARROLLO DEL MODELO HOLÍSTICO DE ANALÍTICA AVANZADA

Este capítulo desarrolla de manera sistemática el Modelo Holístico de Analítica Avanzada (MHAA) conforme a las seis fases del marco de Investigación en Ciencias del Diseño (DSR). El modelo se configura como un artefacto analítico destinado a abordar el problema organizacional de optimizar la toma de decisiones mediante una arquitectura integrada, explícitamente estructurada y metodológicamente fundamentada.

En lugar de privilegiar la contrastación estadística de hipótesis, el capítulo expone —con criterios de rigor y trazabilidad— el proceso de diseño del artefacto: desde la problematización inicial y la definición de objetivos hasta la especificación arquitectónica, la demostración funcional y su validación. Aunque las fases de la DSR se presentan de forma secuencial, su articulación obedece a una lógica interna que procura asegurar coherencia metodológica entre los propósitos investigativos y la arquitectura analítica finalmente configurada.

Se otorga particular atención a la lógica constructiva del modelo —su macroarquitectura, la descomposición funcional y los mecanismos de validación incorporados— con el fin de evidenciar cómo los fundamentos teóricos se traducen en una configuración analítica integrada. La convergencia entre las fases de la DSR y el ciclo clásico de toma de decisiones no opera como mera analogía estructural; más bien, consolida la coherencia interna del modelo propuesto frente a la complejidad inherente a los entornos organizativos.

4.1 Estructuración de las fases del modelo holístico de analítica avanzada para establecer la arquitectura del artefacto analítico.

4.1.1 Identificación del problema y motivación (Fase DSR 1)

Según estudios anteriores, numerosas instituciones enfrentan problemas para tomar decisiones efectivas a causa de prácticas analíticas diseminadas, una integración insuficiente de los estándares de calidad y seguridad de la información, así como

conexiones débiles entre la visualización y los indicadores clave de rendimiento. Estas condiciones tienden a producir procesos de decisión mal organizados, sin soporte técnico y con trazabilidad limitada, lo que disminuye su efectividad en general.

En esta situación, es evidente la necesidad de una solución integral que combine los principios de gobernanza de datos, las habilidades analíticas avanzadas y los métodos de visualización en un solo marco. El propósito no es únicamente brindar resultados analíticos, sino garantizar que sean coherentes, claros y estén en consonancia con las metas de la organización.

Con base en lo anterior, esta investigación se plantea la siguiente pregunta:
¿Cómo el desarrollo de un modelo holístico de analítica avanzada optimiza la toma de decisiones organizacionales?

4.1.2 Definir los objetivos de una solución (Fase DSR 2)

Para enfrentar los retos detectados, esta investigación busca diseñar y validar un modelo holístico que integre las distintas etapas de la analítica—desde la descriptiva hasta la autónoma—junto con prácticas de gobernanza y gestión de datos. No se trata de acumular tecnologías analíticas, sino de proponer una estructura clara que fortalezca la toma de decisiones basada en datos (Data-Driven Decision-Making).

El modelo tiene como meta transformar los datos en conocimiento útil que realmente apoye decisiones estratégicas. En este sentido, funciona como una guía práctica para organizaciones que buscan alinear sus esfuerzos analíticos con los procesos de decisión, más allá de una propuesta teórica o puramente tecnológica.

Este objetivo responde a las diversas limitaciones detectadas en la bibliografía analizada, tales como:

- la fragmentación entre los modelos de análisis y gestión de datos,
- la ausencia de una estructura metodológica que integre la calidad, la seguridad y la gobernanza,
- el uso limitado de la visualización avanzada para interpretar datos críticos,

- y la falta de alineación entre los resultados analíticos y los procesos reales de toma de decisiones en las organizaciones.

Por lo tanto, la solución propuesta tiene como objetivo llenar estos vacíos mediante el desarrollo de un modelo holístico de analítica avanzada para optimizar la toma de decisiones organizacionales que no solo represente el ciclo completo del valor analítico, sino que también sea viable en entornos reales, sea evaluado por el juicio de expertos y esté respaldado por los principios del diseño científico formal. En este sentido, el modelo propuesto busca:

- a) Comprender el contexto organizacional para alinear los objetivos estratégicos con los requerimientos analíticos.
- b) Asegurar la calidad, consistencia y limpieza de los datos como base del proceso analítico.
- c) Integrar datos de múltiples fuentes mediante procesos normalizados y controlados.
- d) Aplicar técnicas de analítica multietapa (descriptiva, predictiva, prescriptiva y autónoma) para generar valor en la toma de decisiones.
- e) Visualizar e interpretar los resultados analíticos mediante herramientas que fomenten la toma de decisiones basadas en evidencia.

Las fases que componen el modelo holístico de analítica avanzada no surgen de manera arbitraria, sino que se derivan de la integración sistemática de marcos teóricos consolidados en la toma de decisiones organizacionales, la gestión de datos y la analítica avanzada. En particular, el modelo articula los principios del proceso de toma de decisiones de Simon (1960) con las buenas prácticas de gestión de datos (DAMA-DMBOK, normas ISO) y la cadena de valor analítica propuesta por Provost y Fawcett (2013). La Tabla 5 presenta la estructura conceptual del modelo, identificando sus fases, su propósito y su sustento teórico.

Tabla 5***Estructura conceptual del Modelo Holístico de Analítica Avanzada***

Fase del modelo holístico	Propósito estructural	Fundamento teórico / marco de referencia
Contexto organizacional	Comprender el entorno organizacional, los objetivos estratégicos, las necesidades decisionales y los requerimientos analíticos antes del tratamiento de datos.	Simon (1960); Davenport & Harris (2007); Provost & Fawcett (2013); DDDM
Limpieza y calidad de datos	Garantizar la confiabilidad, consistencia y precisión de los datos como condición previa para el análisis avanzado.	DAMA-DMBOK; ISO/IEC 25012; ISO 8000; Kimball
Integración de datos	Unificar datos heterogéneos desde múltiples fuentes mediante procesos controlados y semánticamente consistentes.	Inmon; Kimball; ISO/IEC 11179; DAMA-DMBOK
Capacidades de analítica avanzada	Aplicar técnicas analíticas descriptivas, predictivas, prescriptivas y autónomas para generar conocimiento accionable.	Provost & Fawcett (2013); Davenport (2014); CRISP-DM
Visualización de datos	Traducir los resultados analíticos en representaciones comprensibles que faciliten la interpretación y la acción.	Few (2009); Knaflitz (2015); DAMA-DMBOK

A partir de la integración de los marcos teóricos revisados y del análisis sistemático de la literatura, se identificaron cinco fases estructurales que conforman el modelo holístico de analítica avanzada. Estas fases, su propósito, sustento conceptual y rol dentro del proceso de toma de decisiones se presentaron en el cuadro anterior. Sobre esta base conceptual, se diseñó la arquitectura estructurada del modelo holístico, el cual se representa gráficamente en la Figura 6.

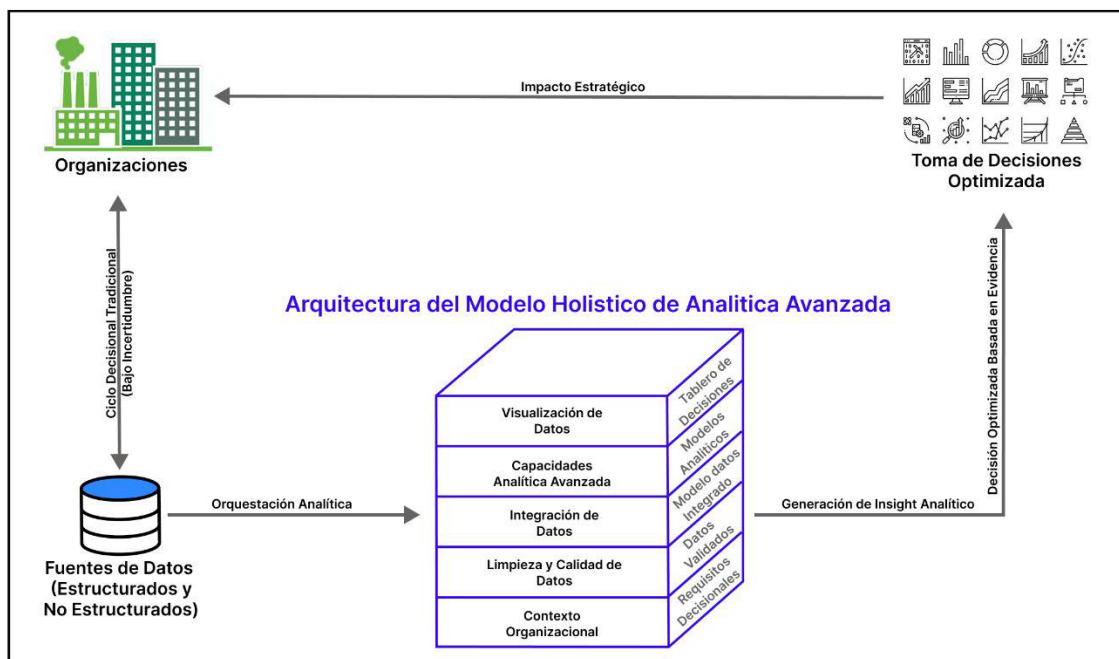


Figura 6 Arquitectura Estructurada del Modelo Holístico de Analítica Avanzada

4.2 Diseño del Modelo Holístico de Analítica Avanzada mediante la descomposición funcional de su arquitectura, especificando componentes y flujos de integración.

4.2.1 Diseño y desarrollo: artefacto (Fase DSR 3)

La representación gráfica del modelo holístico propuesto se muestra en la Figura 2, que detalla, de manera secuencial y funcional, los componentes, flujos y acciones que conforman el artefacto desarrollado.

La Tabla 6 muestra los elementos que conforman la fase de Contexto Organizacional, los cuales son clave y están destinados a establecer las bases estratégicas, los requisitos analíticos y la comprensión del entorno organizacional antes de trabajar con los datos. Estas actividades garantizan que el proceso de análisis avanzado esté alineado con los objetivos comerciales, considerando las necesidades reales y las fuentes de datos disponibles.

Tabla 6
Contexto Organizacional

Actividad	Objetivo	Marcos/ normas de referencia
Planificación estratégica y objetivos.	Definir la visión estratégica y los objetivos para la implementación analítica.	Toma de decisiones basada en datos (DDDM)
Identificación de necesidades y problemas.	Detectar problemas y oportunidades para la intervención analítica.	Modelo de proceso para la toma de decisiones
Requisitos analíticos.	Establecer las capacidades analíticas necesarias para satisfacer las necesidades empresariales.	DAMA-DMBOK
Mapeo inicial de fuentes de datos.	Identificar y enumerar las fuentes de datos internas y externas.	DAMA-DMBOK, ISO/IEC 11179

La Tabla 7 muestra los elementos que conforman la fase de Limpieza y Calidad de Datos para garantizar que dichos datos utilizados sean precisos, coherentes y útiles. Esta etapa se basa en normas como ISO 25012 e ISO 8000, que sirven de guía para la gestión de la calidad de los datos, y emplea técnicas como la elaboración de perfiles y la limpieza, que son esenciales para realizar análisis fiables.

Tabla 7
Limpieza y calidad de los datos

Actividad	Objetivo	Marcos/ Normas de referencia
Perfilado de datos	Analizar los datos para comprender su estructura y calidad.	DAMA-DMBOK, ISO 8000
Evaluación de la calidad	Evaluar la integridad, precisión y coherencia de los datos.	ISO/IEC 25012, DAMA-DMBOK
Reglas de calidad y validación	Definir y aplicar controles de validación de datos.	ISO/IEC 25012, Kimball

Limpieza de datos	Corregir errores, eliminar duplicados y gestionar los valores que faltan.	Crisp-DM, ISO 8000
Normalización y estandarización	Asegurarse de que los datos tengan un formato coherente.	ISO/IEC 11179, Kimball
Consolidación de duplicados	Combinar registros duplicados para garantizar la coherencia.	DAMA-DMBOK
Metadatos de calidad	Documentar las reglas y los atributos de calidad.	ISO/IEC 11179
Supervisión continua de la calidad	Realizar un seguimiento continuo de los indicadores clave de rendimiento (KPI) de la calidad de los datos.	DataOps, ISO 8000

La Tabla 8 muestra cómo se consolidan, transforman y unifican los datos procedentes de múltiples fuentes, en la fase de Integración de Datos. Este proceso se basa en prácticas como las de Kimball e Inmon para la integración y el modelado, y cuenta con el respaldo de herramientas ETL y normas como la ISO/IEC 11179 para la gestión semántica y de metadatos.

Tabla 8
Integración de Datos

Actividad	Objetivo	Marcos/ Normas de referencia
Mapeo semántico y normalización.	Alinear y traducir términos entre fuentes.	ISO/IEC 11179
Transformación y unificación.	Estandarizar formatos y estructuras.	Kimball, Inmon
Modelado de datos integrado.	Crear una representación de datos unificada.	Inmon, DAMA-DMBOK
Resolución de conflictos.	Resolver inconsistencias en los datos.	Kimball
Uso de herramientas ETL.	Extraer, transformar y cargar datos.	Kimball, DAMA-DMBOK
Gestión de metadatos.	Organizar metadatos para su integración y recuperación.	DAMA-DMBOK, ISO/IEC 11179
Control de versiones.	Realizar un seguimiento y	DataOps

	gestionar los cambios en los datos.	
Validación semántica.	Garantizar que los datos integrados cumplan con la semántica definida	ISO/IEC 11179

La Tabla 9 detalla las diferentes Capacidades de la Analítica Avanzada (descriptivas, predictivas, prescriptivas y autónomas), alineadas con la cadena de valor del análisis de datos. Estas capacidades permiten extraer valor de los datos utilizando algoritmos y modelos específicos según el nivel de madurez analítica.

Tabla 9
Capacidades de Analítica Avanzada

Actividad	Objetivo	Marcos/ Normas de referencia
Análisis descriptivo	Comprender el comportamiento pasado mediante resúmenes estadísticos.	Cadena de valor: comprensión del negocio; algoritmos: histogramas, agrupamiento.
Análisis predictivo	Predecir resultados futuros utilizando modelos.	Cadena de valor: minería de datos; algoritmos: regresión, árboles de decisión.
Análisis prescriptivo	Recomendar acciones basadas en predicciones.	Cadena de valor: implementación; algoritmos: optimización, simulación.
Análisis autónomo	Permitir que los sistemas decidan y actúen de forma independiente.	Cadena de valor: automatización; algoritmos: aprendizaje por refuerzo, aprendizaje profundo.

La Tabla 10 describe las actividades necesarias para transformar los resultados analíticos en elementos visuales comprensibles, en consonancia con los principios del análisis visual, la narración de historias y la toma de decisiones basada en datos. Esta fase refuerza la acción organizativa mediante paneles de control, indicadores clave de rendimiento (KPI) y retroalimentación visual continua.

Tabla 10
Visualización de Datos

Actividad	Objetivo	Marcos/ Normas de referencia
Narrativa basada en datos.	Comunicar conocimientos a través de narrativas.	Narrativa para datos, análisis visual
Contextualización con KPI.	Presentar conocimientos alineados con los objetivos de rendimiento.	DAMA-DMBOK
Diseño de paneles de control.	Visualizar métricas clave con claridad.	Análisis visual, Kimball
Alertas y recomendaciones automáticas.	Proporcionar orientación proactiva basada en reglas.	Análisis prescriptivo
Evaluación del impacto en las decisiones.	Evaluar la influencia analítica en los resultados.	Toma de decisiones basada en modelos de procesos
Interacción usuario-sistema.	Permitir a los usuarios explorar e interactuar con los datos.	Análisis visual, diseño centrado en el ser humano
Bucle de retroalimentación.	Incorporar los comentarios de los usuarios y del sistema en la iteración.	DataOps, mejora continua

Los Cuadros precedentes han permitido describir de manera sistemática los componentes, actividades y marcos de referencia asociados a cada una de las fases del modelo holístico de analítica avanzada. En conjunto, estas fases configuran una estructura integrada que articula el contexto organizacional, la gestión de la calidad y la integración de los datos, el desarrollo de capacidades analíticas avanzadas y la visualización orientada a la toma de decisiones. Como resultado de esta construcción conceptual y metodológica, en la Figura 7 se presenta el diseño general del modelo holístico de analítica avanzada, que sintetiza gráficamente la secuencia funcional y la interrelación de sus fases en el proceso de toma de decisiones organizacionales.

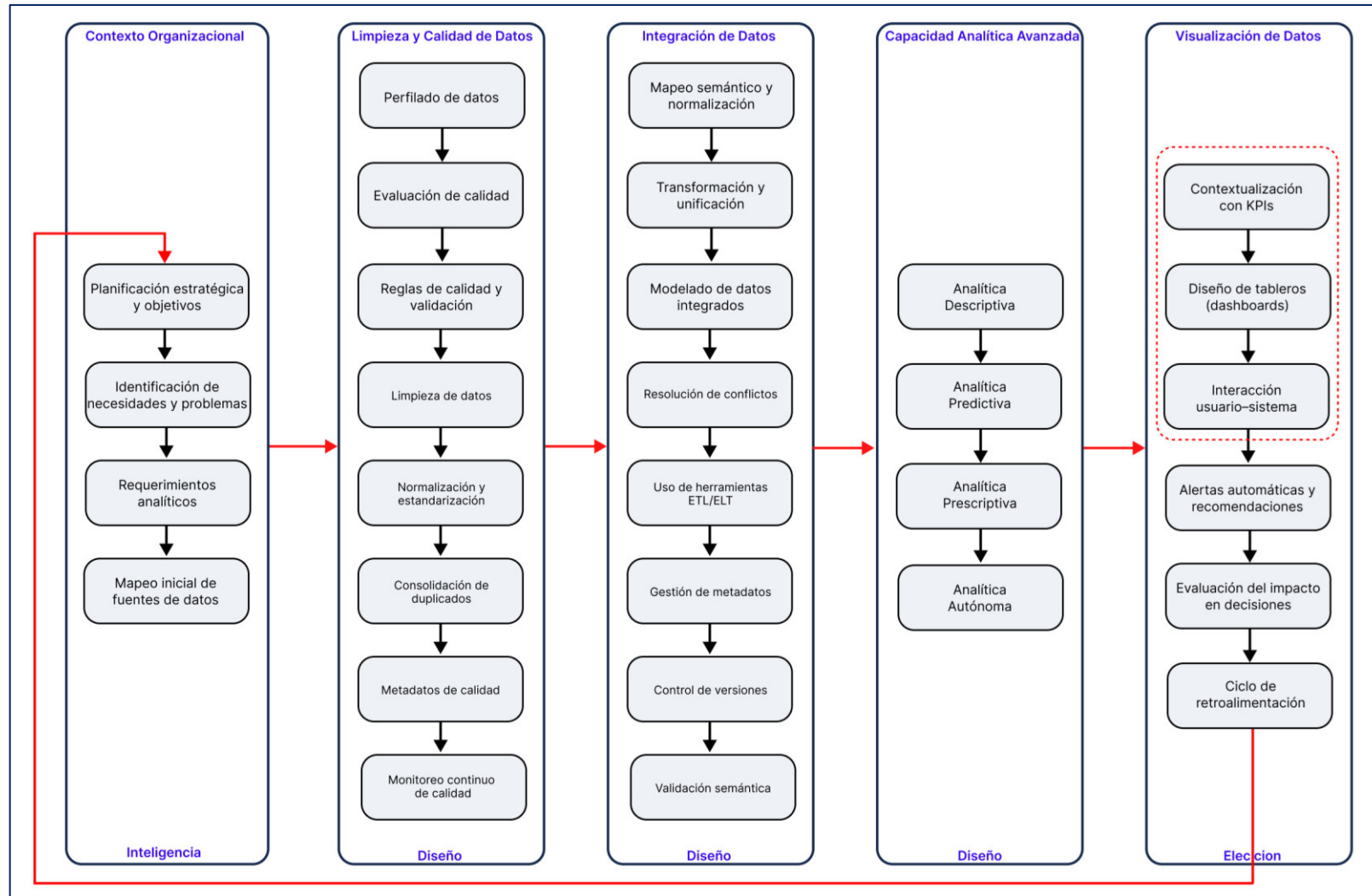


Figura 7 *Modelo Holístico de Analítica Avanzada*

4.2.2 Demostración (Fase DSR 4)

Para demostrar la aplicabilidad del artefacto propuesto, presentamos una alineación entre las fases del modelo holístico de análisis avanzado y las etapas del modelo clásico del proceso de toma de decisiones, que incluye tres momentos clave: inteligencia, diseño y elección. Esta correspondencia, que se muestra en la Tabla 11, nos permite observar cómo el modelo desarrollado ofrece un apoyo estructurado en cada etapa del proceso de toma de decisiones.

Tabla 11

Alineación del modelo holístico y el modelo de toma de decisiones

Fase del Modelo Holístico	Descripción funcional	Fase de Toma de Decisiones
Contexto organizativo	Establece objetivos estratégicos, identifica necesidades, define requisitos analíticos y mapea fuentes de datos.	Inteligencia
Limpieza y calidad de los datos	Garantiza la fiabilidad de los datos mediante la creación de perfiles, la validación, la limpieza, la estandarización y el control de calidad.	Diseño
Integración de datos	Integra y armoniza los datos utilizando herramientas ETL, gestión de metadatos y validación semántica.	Diseño
Capacidades de analíticas avanzadas	Aplica análisis descriptivos, predictivos, prescriptivos y autónomos para explorar alternativas.	Diseño
Visualización de datos	Presenta información analítica mediante narrativas, KPI, paneles de control e interacción con el usuario para apoyar la toma de decisiones.	Elección

La fase inicial del modelo, denominada Contexto organizativo, está diseñada para aplicarse en cualquier tipo de organización, sin importar su tamaño, sector o nivel de

madurez analítica. Su propósito es definir con precisión las necesidades, las fuentes de datos disponibles y la planificación de acciones futuras.

Para validar el modelo, se contó con la participación de siete expertos con experiencia en análisis de datos, visualización y toma de decisiones organizacionales. Los resultados, presentados a continuación, muestran una valoración positiva en cuanto a su utilidad, pertinencia y coherencia metodológica.

4.3 Validación del Modelo Holístico de Analítica Avanzada mediante el juicio de expertos y un caso aplicado

4.3.1 Evaluación (Fase DSR 5)

4.3.1.1 Con Juicios de Expertos

Para evaluar la pertinencia, utilidad y viabilidad del modelo holístico, se aplicó un cuestionario estructurado a un grupo de expertos en análisis avanzado y toma de decisiones. Este instrumento, basado en los criterios metodológicos previamente descritos, tuvo como objetivo recoger información clave sobre el funcionamiento del modelo en contextos organizacionales.

4.3.1.1.1 Interpretación de los resultados de la validación

En la Tabla 12, se muestran los resultados de las preguntas formuladas en el Capítulo IV, específicamente en la Sección 4.6 “Técnicas de recopilación de datos”, cuyos resultados fueron obtenidos a partir de la evaluación realizada por siete expertos en los campos del análisis de datos, la visualización y la inteligencia empresarial. Cada pregunta se calificó en una escala Likert del 1 al 5, donde 1 representa “totalmente en desacuerdo” y 5 representa “totalmente de acuerdo”. Se calcularon tres indicadores para cada pregunta: promedio, desviación estándar y porcentaje de respuestas iguales o superiores a 4. Para la obtención de la información de los expertos se utilizó una herramienta de evaluación propia, diseñada a partir de la experiencia y aportes metodológicos de diversos autores, la cual se describe en el Capítulo IV, Tabla 4.

Tabla 12***Resultados de la validación del modelo según el criterio de los expertos***

Pregunta	Media	Desviación Estándar	Porcentaje ≥ 4
P1	3.71	1.03	57.14%
P2	4.14	1.12	71.43%
P3	3.57	0.49	57.14%
P4	3.71	0.70	57.14%
P5	3.57	1.18	57.14%
P6	3.71	0.88	71.43%
P7	3.71	0.88	71.43%
P8	4.57	0.73	85.71%
P9	3.29	1.03	42.86%
P10	3.71	1.03	57.14%

- a) **Media:** La puntuación media de cada pregunta indica el grado general de aceptación por parte de los expertos con respecto a cada dimensión evaluada en el modelo. La mayoría de las medias se sitúan entre 3,5 y 4,5, lo cual es positivo y muestra una tendencia favorable hacia la utilidad y aplicabilidad del modelo propuesto.
- b) **Desviación estándar:** Este valor mide la variabilidad de las respuestas. Los valores más bajos indican consenso entre los expertos. Por ejemplo, la pregunta Q3 tiene una desviación baja (0,49), lo que sugiere que los expertos están muy de acuerdo en sus respuestas. Por el contrario, una desviación como 1,18 (Q5) indica una mayor dispersión y posiblemente opiniones divergentes sobre esa dimensión.
- c) **Porcentaje ≥ 4 :** Este indicador representa el grado de aceptación fuerte (calificaciones altas). Se puede observar que:
- La pregunta Q8 (visualización e interpretación de los resultados) obtuvo el mayor nivel de aceptación (85,71 %), lo que refleja un alto consenso positivo.

- La pregunta Q9 obtuvo el porcentaje más bajo (42,86 %), lo que indica un área de mejora. Es posible que esta pregunta aborde un aspecto que aún suscita dudas o requiere un refuerzo conceptual.

4.3.1.1.2 Análisis consolidado por criterios de validación

Si bien el análisis por ítem permite comprender el comportamiento individual de cada dimensión evaluada, el objetivo específico de validación de la presente investigación (OE_c) establece la necesidad de evaluar el modelo holístico en términos de pertinencia, utilidad y viabilidad, conforme a los principios de la Design Science Research. En ese sentido, con el fin de responder directamente a dicho objetivo, las preguntas del instrumento fueron agrupadas en estos tres criterios de validación, considerando su afinidad conceptual y funcional. La Tabla 13 presenta los resultados agregados por criterio, permitiendo evaluar de manera sintética el nivel de aceptación del modelo desde una perspectiva integradora.

Tabla 13

Nivel de aceptación del modelo por criterios

Criterio	Media	Desv. Est.	% ≥ 4
Pertinencia	4.00	0.82	67.86%
Viabilidad	3.95	0.80	66.67%
Utilidad	4.24	0.77	80.95%

Los resultados de la validación evidencian que el modelo holístico de analítica avanzada presenta un alto nivel de pertinencia, viabilidad y utilidad. En particular, la pertinencia del modelo alcanza una media de 4.00, lo que confirma su adecuada alineación con el problema de investigación y el dominio organizacional. La viabilidad obtiene una media de 3.95, reflejando que el modelo es factible de implementar y replicar en contextos reales. Finalmente, la utilidad del modelo alcanza una media de 4.24, con más del 80% de valoraciones altas, lo que evidencia su aporte significativo a la optimización de la toma de decisiones organizacionales. Estos resultados confirman el cumplimiento del Objetivo Específico de validación (OE_c), conforme a los principios de evaluación establecidos por la Design Science Research.

4.3.1.2 Evaluación con un Caso Aplicado

Para completar la fase de evaluación del Modelo Holístico de Análisis Avanzado (MHAA) y en consonancia con el marco de la Investigación en Ciencias del Diseño (DSR), en esta sección se presenta un caso aplicado desarrollado para demostrar la eficacia operativa del artefacto propuesto en relación con el Objetivo Específico 3.

En el ámbito de este estudio, la evaluación del modelo se llevó a cabo en dos niveles complementarios. En primer lugar, se evaluaron las dimensiones de relevancia, utilidad y viabilidad mediante el juicio de expertos, como se ha descrito anteriormente. En segundo lugar, el caso aplicado se diseñó para examinar el rendimiento operativo del modelo, entendido como su capacidad para funcionar de manera coherente y estructural cuando se ejecuta con datos del mundo real, apoyando eficazmente los procesos de toma de decisiones de la organización.

Para ello, se seleccionó el conjunto de datos «Registros de servicios prestados a pacientes asegurados - Seguro médico integral (SIS)», que abarca los años 2017, 2018, 2019 y 2020. El conjunto de datos se eligió por su relevancia institucional, continuidad temporal, volumen significativo y diversidad de variables. Estas características permiten la activación y prueba de todas las capas principales del Modelo Holístico de Análisis Avanzado sin introducir restricciones artificiales o simplificaciones excesivas.

En consecuencia, el caso aplicado constituye una prueba empírica de funcionamiento del modelo, complementaria a la validación conceptual realizada mediante juicio de expertos, y permite cerrar de manera coherente la fase de evaluación del enfoque DSR adoptado en esta investigación. A continuación, se presentan los resultados obtenidos por cada fase de la estructura del modelo holístico propuesto.

4.3.1.2.1 Contexto Organizacional

La primera fase del modelo holístico de analítica avanzada corresponde al contexto organizacional, cuyo objetivo es establecer el marco estratégico y operativo sobre el cual se desarrolla el proceso analítico. En esta fase se consideran la planificación estratégica, la identificación de necesidades y problemas, la definición de requerimientos

analíticos y el mapeo inicial de las fuentes de datos.

En el contexto del presente caso aplicado, el problema organizacional se enmarca en la necesidad de optimizar la toma de decisiones en el sistema público de salud, particularmente en la gestión de la demanda de atenciones del Seguro Integral de Salud (SIS). La alta variabilidad en la cantidad de atenciones por región, servicio, grupo etario y periodo temporal genera escenarios complejos que requieren el uso de enfoques analíticos avanzados para su adecuada comprensión y gestión.

Como parte de la planificación analítica, se definió como objetivo principal evaluar la capacidad del modelo holístico para transformar datos reales del sector salud en información estructurada, analíticamente consistente y útil para la toma de decisiones, alineándose directamente con el objetivo específico c) de la investigación.

El mapeo inicial de las fuentes de datos se realizó a partir del repositorio oficial Datos Abiertos del Estado Peruano, específicamente del portal datosabiertos.gob.pe, donde se encuentra publicado el dataset del SIS (Ministerio de Salud del Perú [MINSA], 2017). Esta fuente fue seleccionada por su carácter público, trazabilidad institucional y cumplimiento de los lineamientos de gobernanza de datos del Estado peruano, lo que garantiza la confiabilidad y legitimidad de la información utilizada en el análisis.

De este modo, la fase de contexto organizacional establece las bases conceptuales y operativas necesarias para la aplicación ordenada y reproducible de las siguientes fases del modelo holístico de analítica avanzada, asegurando la coherencia entre el problema organizacional, los datos disponibles y los objetivos analíticos planteados.

4.3.1.2.2 Limpieza y Calidad de Datos

La Tabla 14 evidencia un alto nivel de completitud del dataset. Las variables críticas para el análisis (tiempo, ubicación, servicio, sexo y métrica de atenciones) presentan ausencia total de valores nulos, lo que confirma la idoneidad del conjunto de datos para procesos analíticos posteriores. Las bajas tasas de nulos detectadas en variables territoriales no comprometen la calidad general del análisis

Tabla 14***Perfil de calidad del dataset antes del proceso de limpieza***

Variable	Tipo	Nº Filas	% Nulos	Nº Valores únicos	Observación
Provincia	Texto	5,432,903	0.051%	196	Baja incompletitud
Distrito	Texto	5,432,903	0.051%	1,717	Baja incompletitud
Año	Entero	5,432,903	0.000%	4	Completo
Anio_Archivo	Entero	5,432,903	0.000%	4	Trazabilidad correcta
Sexo	Texto	5,432,903	0.000%	2	Dominio consistente
Grupo_Edad	Texto	5,432,903	0.000%	6	Dominio consistente
Mes	Entero	5,432,903	0.000%	12	Rango válido
Region	Texto	5,432,903	0.000%	26	Completo
Cod_Servicio	Texto	5,432,903	0.000%	62	Completo
Atenciones	Entero	5,432,903	0.000%	1,246	Variable métrica

La Tabla 15 confirma que el dataset no presenta violaciones estructurales en variables temporales ni métricas.

Esto indica que los datos provienen de un proceso institucional formal, con controles previos, reforzando la confiabilidad del insumo analítico.

Tabla 15***Resultados de las reglas de calidad (antes de la limpieza)***

Regla de validación	Resultado
Años fuera del rango definido	0
Meses fuera del rango 1–12	0
Atenciones negativas	0
Inconsistencias servicio–código	0
Dominio de sexo	Femenino / Masculino

En la Tabla 16 se identificó un volumen significativo de duplicados al nivel del grano analítico, lo cual es habitual en datasets administrativos agregados por múltiples dimensiones. La consolidación permitió preservar la métrica de atenciones mediante agregación, sin pérdida de información, optimizando el dataset para análisis predictivo y

visualización. Este proceso mejora la eficiencia analítica y reduce el ruido estructural del conjunto de datos.

Tabla 16

Detección de duplicados y consolidación analítica

Indicador	Valor
Registros iniciales	5,432,903
Duplicados detectados (grano analítico)	500,561
Registros después de consolidación	4,932,342
Porcentaje de reducción	9.21%

En la Tabla 17 el perfil posterior confirma que el proceso de limpieza eliminó completamente duplicados analíticos, manteniendo la integridad y consistencia del dataset.

El conjunto de datos resultante cumple con los criterios de calidad, validez y consistencia, quedando preparado para las fases de integración y analítica avanzada del modelo.

Tabla 17

Perfil de calidad del dataset después del proceso de limpieza

Indicador	Antes	Después
Registros totales	5,432,903	4,932,342
% promedio de valores nulos	0.0%	0.0%
Duplicados al grano analítico	500,561	0
Atenciones negativas	0	0
Meses inválidos	0	0
Años inválidos	0	0

Los valores extremos identificados corresponden a escenarios reales de alta demanda de atención, propios del dominio de salud. Por tanto, no se eliminaron, sino que fueron documentados como parte del control de calidad, como se muestra en la Tabla 18, evitando distorsiones en el análisis predictivo posterior.

Tabla 18***Distribución y control de valores extremos (outliers)***

Indicador	Antes	Después
Observaciones fuera del rango P1–P99	54,065	48,191
Percentil inferior (P1)	1	1
Percentil superior (P99)	68	75

La Fase 2 permitió transformar el dataset integrado en un conjunto de datos analíticamente consistente, mediante perfilado, validación, limpieza y consolidación. Los resultados evidencian un alto nivel de calidad de los datos, con eliminación total de duplicados al nivel analítico y preservación de la métrica de interés. Este proceso garantiza la confiabilidad del dataset para las fases posteriores de integración, analítica avanzada y visualización, demostrando la aplicabilidad operativa del modelo holístico propuesto.

4.3.1.2.3 Integración de Datos

En la Tabla 19 se evidencia como la integración de datos permitió transformar un dataset operativo de gran volumen en una vista analítica agregada, reduciendo significativamente el número de registros sin pérdida de información relevante. La reducción del 89.95% responde a la definición de un grano decisonal, adecuado para análisis predictivo y soporte a decisiones, garantizando consistencia temporal, territorial y funcional. La ausencia de conflictos temporales y métricos evidencia una integración estructuralmente sólida.

Tabla 19***KPIs del proceso de integración de datos***

Indicador	Valor
Registros de entrada (df_clean)	4,932,342
Columnas de entrada	14
Registros post-unificación	4,932,342

Indicador	Valor
Registros integrados (vista analítica)	495,677
Columnas integradas	10
Reducción de registros (%)	89.95%
Grano analítico aplicado	Año, Mes, Región, Provincia, Servicio, Sexo, Grupo de Edad
Conflictos de año detectados	0
Meses inválidos detectados	0
Atenciones nulas detectadas	0
Atenciones negativas detectadas	0

En la Tabla 20 las reglas semánticas confirman que el dataset integrado cumple con las restricciones de negocio definidas para el dominio de salud.

La inexistencia de violaciones demuestra que la integración no introdujo inconsistencias lógicas, reforzando la confiabilidad del conjunto de datos para análisis avanzado y visualización.

Tabla 20

Resultados de la validación semántica del dataset integrado

Regla de negocio	Violaciones detectadas
Atenciones negativas	0
Sexo fuera de dominio	0
Año fuera del rango 2017–2020	0
Mes fuera del rango 1–12	0

En la Tabla 21 el resumen de dominios evidencia una integración semántica efectiva, consolidando catálogos homogéneos a partir de múltiples fuentes temporales. Esta estandarización permite comparabilidad interanual y territorial, facilitando análisis transversales y la construcción de indicadores consistentes.

Tabla 21***Resumen de dominios y catálogos integrados***

Dominio	N° de categorías
Región	26
Provincia	197
Servicio	62
Sexo	2
Grupo de edad	6

La Tabla 22 muestra la estructura resultante corresponde a un modelo de hechos agregados, optimizado para análisis descriptivo, predictivo y prescriptivo. La inclusión de llaves de integración permite trazabilidad y facilita la reutilización del dataset en procesos analíticos y visualizaciones posteriores.

Tabla 22***Estructura del dataset integrado (diccionario resumido)***

Variable	Tipo	Descripción
ANIO	Entero	Año de atención
MES	Entero	Mes de atención
REGION	Texto	Región administrativa
PROVINCIA	Texto	Provincia administrativa
SERVICIO	Texto	Tipo de servicio de salud
SEXO	Texto	Sexo del paciente
GRUPO_EDAD	Texto	Rango etario
ATENCIONES	Entero	Número de atenciones
KEY_RAW	Texto	Llave analítica legible
KEY_INTEGRACION	Texto	Identificador único integrado

La Fase 3 permitió integrar y unificar el conjunto de datos limpio en una vista analítica consistente, mediante mapeo semántico, normalización de dominios y definición de un grano decisonal adecuado. El proceso redujo significativamente la dimensionalidad del dataset sin pérdida de información relevante, garantizando consistencia estructural y semántica. Como resultado, se obtuvo un dataset integrado,

trazable y validado, apto para la aplicación de técnicas de analítica avanzada y visualización orientada a la toma de decisiones.

4.3.1.2.4 Capacidad Analítica Avanzada

a) Analítica Descriptiva

La Tabla 23 presenta los indicadores descriptivos globales del caso aplicado, los cuales permiten dimensionar la magnitud y cobertura del sistema analizado. El elevado volumen de atenciones evidencia la relevancia operativa del sistema de salud, mientras que el número de regiones y servicios confirma su alcance nacional y funcional. Estos indicadores establecen la línea base cuantitativa sobre la cual se desarrollan los análisis predictivos y prescriptivos posteriores.

Tabla 23
Indicadores descriptivos globales del sistema de atenciones

Indicador	Valor
Periodo de análisis	Enero 2020 – Diciembre 2020
Total de atenciones	39,753,227
Promedio de atenciones por registro	80.20
Registros analíticos	495,677
Número de regiones	26
Número de servicios	62

La Tabla 24 revela un quiebre estructural significativo durante el segundo trimestre del año 2020, caracterizado por una caída abrupta en el número de atenciones, particularmente en abril. Este comportamiento se asocia a un factor exógeno, correspondiente al inicio de la emergencia sanitaria por COVID-19 y las restricciones operativas impuestas al sistema de salud. La posterior recuperación progresiva evidencia la capacidad de adaptación del sistema. La correcta captura de este patrón confirma que el modelo refleja adecuadamente variaciones contextuales reales, fortaleciendo su validez, descriptiva.

Tabla 24
Tendencia Mensual año 2020

Mes	Atenciones	Variación porcentual mensual (%)
Enero	6,523,844	–
Febrero	6,470,075	-0.82
Marzo	4,224,275	-34.71
Abril	1,230,967	-70.86
Mayo	1,453,957	18.12
Junio	1,893,637	30.24
Julio	2,229,626	17.74
Agosto	2,380,042	6.75
Septiembre	2,841,108	19.37
Octubre	3,405,578	19.87
Noviembre	3,621,526	6.34
Diciembre	3,478,592	-3.95

La representación gráfica de la Figura 8 refuerza visualmente el comportamiento identificado En el Cuadro anterior, permitiendo observar con claridad la caída abrupta y la recuperación gradual del sistema. Este tipo de visualización facilita la comprensión de tendencias temporales y sirve como insumo clave para la fase predictiva.

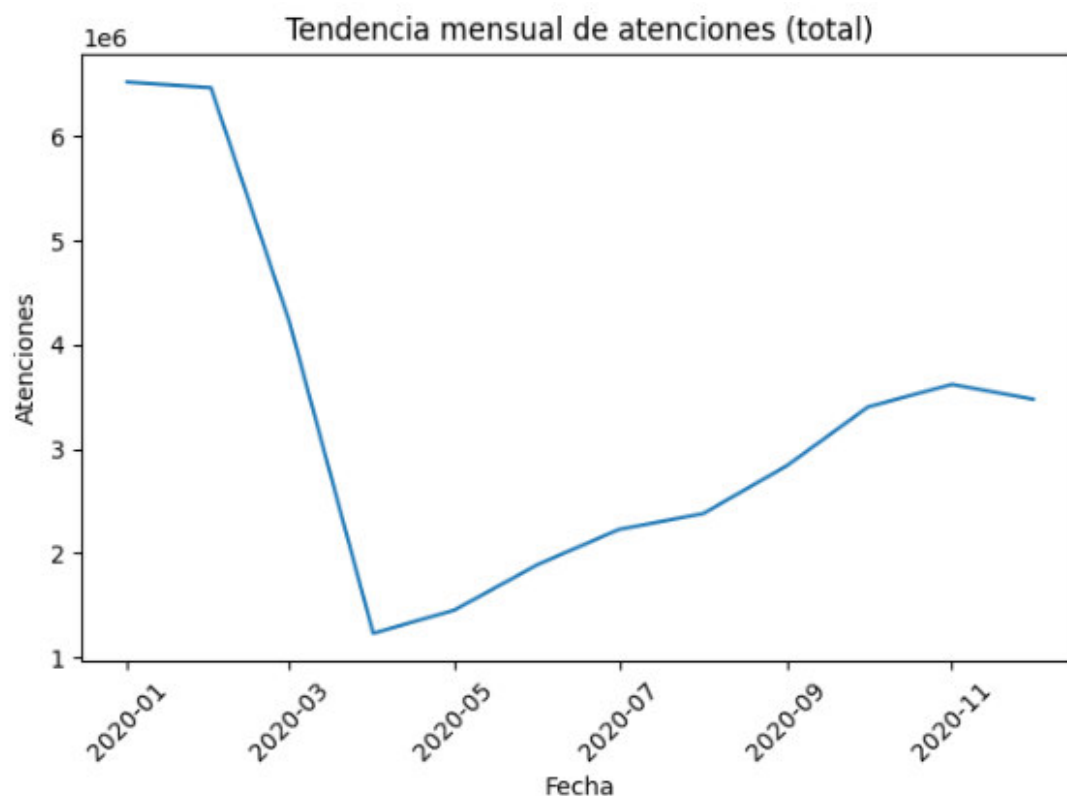


Figura 8 *Evolución mensual del total de atenciones de salud (2020)*

La distribución etaria presentada en la Tabla 25 muestra una alta concentración de atenciones en los extremos de la pirámide poblacional, especialmente en niños menores de cinco años y adultos en edad productiva. Estos resultados permiten identificar segmentos prioritarios para la planificación de recursos y sustentan la segmentación utilizada en las etapas analíticas posteriores.

Tabla 25

Concentración de atenciones

Grupo de edad	Total de atenciones
00 – 04 años	12,118,621
30 – 59 años	9,984,666
18 – 29 años	7,053,570
60+ años	4,376,747
05 – 11 años	3,364,424
12 – 17 años	2,855,199

La analítica descriptiva permitió caracterizar el comportamiento histórico de las

atenciones de salud, identificando tendencias temporales, quiebres estructurales y patrones de segmentación relevantes. Los resultados establecen una línea base sólida y contextualizada que sustenta la aplicación de modelos predictivos y prescriptivos, evidenciando la capacidad del modelo propuesto para analizar datos reales de gran escala y complejidad.

b) Analítica Predictiva

La Tabla 26 presenta el desempeño del modelo predictivo entrenado mediante una partición temporal, utilizando los datos históricos del período 2017–2019 y evaluando su capacidad de generalización sobre el año 2020. El MAE obtenido (109.43) indica que, en promedio, el modelo presenta un error absoluto cercano a 109 atenciones por combinación analítica, mientras que el RMSE (697.66) evidencia la presencia de errores mayores en escenarios de alta demanda, penalizando desviaciones significativas. Estos resultados reflejan un comportamiento esperado en datasets de naturaleza altamente heterogénea y confirman la viabilidad del enfoque predictivo propuesto.

Tabla 26

Métricas del modelo predictivo

Indicador	Valor
Tipo de partición	Temporal (2017–2019 vs 2020)
MAE	109.43
RMSE	697.66
Registros de entrenamiento	1,659,242
Registros de prueba	459,346
Variables categóricas	Región, Servicio, Sexo, Grupo etario
Variable temporal	Índice mensual
Servicios considerados	Top 40
Provincia incluida	No

La Tabla 27 presenta una muestra de los casos con mayor error absoluto entre los valores reales y los predichos por el modelo. Se observa que los errores más elevados se concentran principalmente en combinaciones de alta demanda correspondientes a Lima Metropolitana y al servicio de Consulta Externa. Este comportamiento sugiere la presencia de picos de demanda atípicos y variaciones estructurales que no pueden ser completamente capturadas por un modelo predictivo basado en agregaciones temporales, lo cual es consistente con la literatura sobre analítica predictiva aplicada a servicios de salud.

Tabla 27

Comparación Real vs Predicho (casos con mayor error)

Region	Servicio	Sexo	GrupoEdad	Mes	Real	Predicho	Error abs.
Lima Metropolitana	Consulta Externa	F	30 - 59 Anios	39	9398	116130	106732
Lima Metropolitana	Consulta Externa	F	30 - 59 Anios	40	12184	116130	103946
Lima Metropolitana	Consulta Externa	F	30 - 59 Anios	41	17426	116130	98704
Lima Metropolitana	Consulta Externa	F	30 - 59 Anios	42	25185	116130	90945
Lima Metropolitana	Consulta Externa	F	30 - 59 Anios	43	29260	116130	86870
Lima Metropolitana	Consulta Externa	F	30 - 59 Anios	44	32262	116130	83868
Lima Metropolitana	Consulta Externa	F	30 - 59 Anios	45	40182	116130	75948
Lima Metropolitana	Consulta Externa	F	30 - 59 Anios	47	42207	116130	73923
Lima Metropolitana	Consulta Externa	F	30 - 59 Anios	46	43032	116130	73098
Lima Metropolitana	Consulta Externa	F	60 - Mas Anios	39	4873	75202	70329
Lima Metropolitana	Consulta Externa	F	60 - Mas Anios	40	6223	75202	68979
Lima Metropolitana	Consulta Externa	F	60 - Mas Anios	41	8591	75202	66611

Lima	Consulta	F	60 - Mas	42	11940	75202	63262
Metropolitana	Externa		Anios				
Lima	Consulta	F	60 - Mas	43	13361	75202	61841
Metropolitana	Externa		Anios				
Lima	Consulta	F	60 - Mas	44	15800	75202	59402
Metropolitana	Externa		Anios				
Lima	Consulta	F	60 - Mas	45	21226	75202	53976
Metropolitana	Externa		Anios				
Lima	Consulta	F	60 - Mas	46	22894	75202	52308
Metropolitana	Externa		Anios				
Lima	Consulta	F	60 - Mas	47	23512	75202	51690
Metropolitana	Externa		Anios				
Lima	Consulta	F	30 - 59 Anios	38	68043	116130	48087
Metropolitana	Externa						
Lima	Consulta	M	00 - 04 Anios	39	1326	46157	44831
Metropolitana	Externa						

La configuración del modelo de la Tabla 28 se diseñó considerando la alta cardinalidad de las variables categóricas y el volumen del dataset original. Para garantizar viabilidad computacional sin afectar la validez analítica, se aplicó una estrategia de reducción dimensional basada en la selección de los servicios más frecuentes y la exclusión de variables con alta dispersión. Esta decisión permitió entrenar un modelo robusto y escalable, manteniendo interpretabilidad y alineación con los objetivos del estudio.

Tabla 28

Configuración analítica del modelo

Elemento	Descripción
Tipo de modelo	Random Forest Regressor
Horizonte de predicción	Mensual
Variable objetivo	Número de atenciones
Nivel de agregación	Región–Servicio–Sexo–Grupo etario–Mes
Estrategia de reducción	Top servicios + exclusión de provincia
Estrategia temporal	Índice mensual

La analítica predictiva permitió estimar el comportamiento futuro de la demanda de

La Tabla 29 presenta los **30 escenarios con mayor demanda proyectada**, clasificados como Prioridad ALTA por el modelo prescriptivo. Se observa una clara concentración de escenarios críticos en los servicios de **Consulta Externa** y **Control de Crecimiento y Desarrollo en Menores**, lo cual evidencia áreas de alto impacto operativo para el sistema de salud.

Desde una perspectiva territorial, regiones como **Cajamarca, La Libertad, Cusco, Huánuco y Puno** aparecen recurrentemente, indicando una presión asistencial futura significativa. En términos poblacionales, destacan principalmente los grupos etarios **00–04 Años** y **30–59 Años**, lo que sugiere la necesidad de reforzar tanto la atención preventiva infantil como la atención ambulatoria en adultos.

Las recomendaciones generadas permiten transformar la predicción en **acciones concretas**, orientando la priorización de recursos humanos, logísticos y presupuestales de manera anticipada, reduciendo el riesgo de saturación del sistema.

Recomendaciones prescriptivas (Top 30 – Prioridad ALTA)

Nº	Región	Servicio	Sexo	GrupoEdad	Month Index	Atenciones Pred.	Prioridad	Recomendación
----	--------	----------	------	-----------	----------------	---------------------	-----------	---------------

							Priorizar recursos para Consulta Externa en Cajamarca (Sexo: Femenino, Grupo: 30–59 Anios).
1	Cajamarca	Consulta Externa	Femenino	30–59 Anios	48	1886.75 ALTA	Priorizar recursos para Control de Crecimiento y Desarrollo en Menores en Cajamarca (Sexo: Masculino, Grupo: 00–04 Anios).
2	Cajamarca	Control de Crecimiento y Desarrollo en Menores	Masculino	00–04 Anios	48	1760.17 ALTA	Priorizar recursos para Control de Crecimiento y Desarrollo en Menores en Cajamarca (Sexo: Femenino, Grupo: 00–04 Anios).
3	Cajamarca	Control de Crecimiento y Desarrollo en Menores	Femenino	00–04 Anios	48	1705.34 ALTA	Priorizar recursos para Consulta Externa en La Libertad (Sexo: Femenino, Grupo: 30–59 Anios).
4	La Libertad	Consulta Externa	Femenino	30–59 Anios	48	1678.48 ALTA	Priorizar recursos para Consulta Externa en Cusco (Sexo: Femenino, Grupo: 30–59 Anios).
5	Cusco	Consulta Externa	Femenino	30–59 Anios	48	1380.21 ALTA	

6	Huánuco	Consulta Externa	Femenino	30–59 Anios	48	1331.5	ALTA	Priorizar recursos para Consulta Externa en Huánuco (Sexo: Femenino, Grupo: 30–59 Anios).
7	Cusco	Control de Crecimiento y Desarrollo en Menores	Femenino	00–04 Anios	48	1300.24	ALTA	Priorizar recursos para Control de Crecimiento y Desarrollo en Menores en Cusco (Sexo: Femenino, Grupo: 00–04 Anios).
8	La Libertad	Consulta Externa	Femenino	60+ Anios	48	1257.35	ALTA	Priorizar recursos para Consulta Externa en La Libertad (Sexo: Femenino, Grupo: 60+ Anios).
9	Áncash	Consulta Externa	Femenino	30–59 Anios	48	1056.85	ALTA	Priorizar recursos para Consulta Externa en Áncash (Sexo: Femenino, Grupo: 30–59 Anios).
10	Huánuco	Control de Crecimiento y Desarrollo en Menores	Masculino	00–04 Anios	48	1023.55	ALTA	Priorizar recursos para Control de Crecimiento y Desarrollo en Menores en Huánuco (Sexo: Masculino, Grupo: 00–04 Anios).

11	Cajamarca	Salud Reproductiva (Planificación Familiar)	Femenino	30–59 Anios	48	1020.99	ALTA	Priorizar recursos para Salud Reproductiva en Cajamarca (Sexo: Femenino, Grupo: 30–59 Anios).
12	Puno	Control de Crecimiento y Desarrollo en Menores	Masculino	00–04 Anios	48	888.76	ALTA	Priorizar recursos para Control de Crecimiento y Desarrollo en Menores en Puno (Sexo: Masculino, Grupo: 00–04 Anios).
13	La Libertad	Consulta Externa	Masculino	60+ Anios	48	878.49	ALTA	Priorizar recursos para Consulta Externa en La Libertad (Sexo: Masculino, Grupo: 60+ Anios).
14	Puno	Control de Crecimiento y Desarrollo en Menores	Femenino	00–04 Anios	48	872.99	ALTA	Priorizar recursos para Control de Crecimiento y Desarrollo en Menores en Puno (Sexo: Femenino, Grupo: 00–04 Anios).
15	Cajamarca	Estimulación Temprana para Menores de 36 Meses	Masculino	00–04 Anios	48	859.2	ALTA	Priorizar recursos para Estimulación Temprana en Cajamarca (Sexo: Masculino, Grupo: 00–04 Anios).

							Priorizar recursos para Consulta Externa en Cajamarca (Sexo: Femenino, Grupo: 60+ Anios).
16	Cajamarca	Consulta Externa	Femenino	60+ Anios	48	850.13 ALTA	Priorizar recursos para Estimulación Temprana en Cajamarca (Sexo: Femenino, Grupo: 00–04 Anios).
17	Cajamarca	Estimulación Temprana para Menores de 36 Meses	Femenino	00–04 Anios	48	810.16 ALTA	Priorizar recursos para Salud Reproductiva en Cajamarca (Sexo: Femenino, Grupo: 18–29 Anios).
18	Cajamarca	Salud Reproductiva (Planificación Familiar)	Femenino	18–29 Anios	48	758.33 ALTA	Priorizar recursos para Consulta Externa en Puno (Sexo: Femenino, Grupo: 30–59 Anios).
19	Puno	Consulta Externa	Femenino	30–59 Anios	48	750.93 ALTA	Priorizar recursos para Consulta Externa en Cusco (Sexo: Masculino, Grupo: 00–04 Anios).
20	Cusco	Consulta Externa	Masculino	00–04 Anios	48	744.96 ALTA	Priorizar recursos para Consulta Externa en Cajamarca (Sexo: Femenino,
21	Cajamarca	Consulta Externa	Femenino	18–29 Anios	48	742.54 ALTA	

Grupo: 18–29
Años).

22	Cajamarca	Detección de Problemas en Salud Mental	Femenino	00–04 Años	48	740.22	ALTA	Priorizar recursos para Detección de Problemas en Salud Mental en Cajamarca (Sexo: Femenino, Grupo: 00–04 Años).
23	Cusco	Consulta Externa	Femenino	00–04 Años	48	732.04	ALTA	Priorizar recursos para Consulta Externa en Cusco (Sexo: Femenino, Grupo: 00–04 Años).
24	Cajamarca	Consulta Externa	Masculino	00–04 Años	48	727.97	ALTA	Priorizar recursos para Consulta Externa en Cajamarca (Sexo: Masculino, Grupo: 00–04 Años).
25	La Libertad	Consulta Externa	Masculino	00–04 Años	48	723.78	ALTA	Priorizar recursos para Consulta Externa en La Libertad (Sexo: Masculino, Grupo: 00–04 Años).
26	Huánuco	Consulta Externa	Masculino	00–04 Años	48	721.08	ALTA	Priorizar recursos para Consulta Externa en Huánuco (Sexo: Masculino, Grupo: 00–04 Años).

							Priorizar recursos para Consulta Externa en Cajamarca (Sexo: Masculino, Grupo: 30–59 Anios).
27	Cajamarca	Consulta Externa	Masculino	30–59 Anios	48	715.64 ALTA	Priorizar recursos para Consulta Externa en Cusco (Sexo: Femenino, Grupo: 18–29 Anios).
28	Cusco	Consulta Externa	Femenino	18–29 Anios	48	708.31 ALTA	Priorizar recursos para Consulta Externa en La Libertad (Sexo: Femenino, Grupo: 18–29 Anios).
29	La Libertad	Consulta Externa	Femenino	18–29 Anios	48	701.02 ALTA	Priorizar recursos para Control de Crecimiento y Desarrollo en Menores en Áncash (Sexo: Femenino, Grupo: 00–04 Anios).
30	Áncash	Control de Crecimiento y Desarrollo en Menores	Femenino	00–04 Anios	48	695.87 ALTA	

La Tabla 30 resume el comportamiento global del análisis prescriptivo. A partir de 800 escenarios evaluados, se establecieron umbrales basados en percentiles estadísticos (P75 y P90), permitiendo clasificar la demanda futura en tres niveles de prioridad.

Solo el 10 % de los escenarios fue clasificado como Prioridad ALTA, lo que demuestra que una fracción reducida de combinaciones concentra el mayor riesgo de saturación del sistema. Este resultado es clave para la optimización de recursos, ya que evita intervenciones generalizadas y favorece una focalización estratégica.

El uso de los 40 servicios más representativos garantiza que el análisis se mantenga computacionalmente eficiente sin perder representatividad del sistema real.

Tabla 30

Resumen de prioridades prescriptivas

Indicador	Valor
Escenarios evaluados	800
Umbral P75	181.12
Umbral P90	420.45
Prioridad Alta	80
Prioridad Media	120
Prioridad Baja	600
Month Index	48
Top servicios usados	40
Variables base	Región, Servicio, Sexo, Grupo Edad

La Tabla 31 muestra los diez escenarios con mayor demanda proyectada confirman la concentración del riesgo operativo en determinados servicios y segmentos poblacionales. La reiterada presencia de Consulta Externa y atención infantil refuerza la necesidad de acciones inmediatas en estos frentes.

Este cuadro funciona como un indicador prescriptivo clave, permitiendo a los gestores identificar rápidamente los puntos críticos donde una falta de intervención podría generar colapsos en la atención.

Tabla 31

Escenarios priorizados según demanda estimada de atenciones

Región	Servicio	Sexo	Grupo Edad	Mes	Atenciones Pred.	Prioridad
Cajamarca	Consulta Externa	Femenino	30–59 Años	48	1886.75	ALTA

Cajamarca	Control de Crecimiento y Desarrollo en Menores	Masculino	00–04 Años	48	1760.17	ALTA
Cajamarca	Control de Crecimiento y Desarrollo en Menores	Femenino	00–04 Años	48	1705.34	ALTA
La Libertad	Consulta Externa	Femenino	30–59 Años	48	1678.48	ALTA
Cusco	Consulta Externa	Femenino	30–59 Años	48	1380.21	ALTA
Huánuco	Consulta Externa	Femenino	30–59 Años	48	1331.5	ALTA
Cusco	Control de Crecimiento y Desarrollo en Menores	Femenino	00–04 Años	48	1300.24	ALTA
La Libertad	Consulta Externa	Femenino	60+ Años	48	1257.35	ALTA
Áncash	Consulta Externa	Femenino	30–59 Años	48	1056.85	ALTA
Puno	Control de Crecimiento y Desarrollo en Menores	Masculino	00–04 Años	48	888.76	ALTA

La analítica prescriptiva permitió operacionalizar los resultados de las fases descriptiva y predictiva, transformándolos en recomendaciones priorizadas y accionables, orientadas a la toma de decisiones organizacionales informadas. A través de este enfoque, el modelo desarrollado:

- Traduce la predicción de demanda en escenarios concretos de priorización operativa.
- Permite anticipar contextos críticos de presión asistencial, facilitando la planificación proactiva.
- Optimiza la asignación de recursos sanitarios en función de escenarios futuros estimados.

- Reduce la incertidumbre en la toma de decisiones organizacionales, al basarse en evidencia analítica.
- Demuestra la viabilidad práctica del modelo holístico de analítica avanzada propuesto en la investigación.

Es importante aclarar que, aunque la analítica prescriptiva representa una etapa avanzada y valiosa dentro del proceso, el cierre completo del ciclo de la analítica avanzada se alcanza con la analítica autónoma. Esta última se caracteriza por la capacidad de ejecutar decisiones de forma automática, sin necesidad de intervención humana directa.

Este estudio no contempla la implementación de analítica autónoma, ya que ello implicaría contar con una infraestructura tecnológica más robusta, integración con sistemas en funcionamiento y un entorno organizacional controlado, elementos que van más allá del alcance de esta investigación. No obstante, los resultados obtenidos constituyen un punto de partida confiable y bien fundamentado para explorar, en el futuro, la adopción de modelos de analítica autónoma.

4.3.1.2.5 Visualización

La visualización ocupa un lugar transversal en el modelo holístico de analítica avanzada, en la medida en que articula la representación y comunicación de resultados a lo largo de los niveles descriptivo, predictivo y prescriptivo. No se limita, por tanto, a una función meramente gráfica; más bien, opera como un dispositivo cognitivo que traduce configuraciones de datos complejas en conocimiento interpretable, susceptible de validación y aplicable en entornos organizacionales concretos.

En el plano descriptivo, la visualización sostiene el examen del comportamiento histórico, la identificación de tendencias y la delimitación de segmentos estratégicos. Cuando el análisis se desplaza hacia la capa predictiva, estas representaciones permiten contrastar valores observados y estimados, así como hacer visible la incertidumbre inherente a los modelos. Finalmente, en el nivel prescriptivo, la visualización contribuye a jerarquizar escenarios y explicitar recomendaciones, facilitando decisiones relativas a la asignación de recursos y a la planificación operativa.

En resumen, la visualización no se trata como una fase separada, sino como un componente integrador que conecta las capacidades analíticas y garantiza la coherencia y la utilidad práctica de los resultados.

La figura 9 presenta el dashboard descriptivo, que ofrece una visión consolidada del sistema de servicios sanitarios del SIS entre 2017 y 2020. El indicador principal, que muestra más de 264 millones de servicios sanitarios, refleja la escala nacional del sistema y proporciona contexto para los demás análisis visuales.

- **Atenciones por región (Top 5):**

El gráfico de barras revela una concentración significativa de atenciones en Lima Metropolitana, muy por encima del resto de las regiones. Le siguen, aunque a bastante distancia, Cajamarca, Cusco, Áncash y La Libertad. Este patrón pone en evidencia una fuerte centralización geográfica de la demanda de servicios de salud, influida tanto por la densidad poblacional como por la presencia de infraestructura especializada en la capital. Esta situación plantea desafíos importantes en cuanto a equidad territorial y al objetivo de descentralizar el sistema de salud.

- **Atenciones por servicio (Top 5):**

El gráfico circular destaca que el servicio de Control de Crecimiento y Desarrollo concentra más de la mitad del total de atenciones, seguido por otros servicios orientados a la prevención y atención primaria. Este resultado refuerza el enfoque estratégico del SIS en la prevención y el monitoreo continuo de la salud, especialmente en población infantil, lo cual justifica que estos servicios sean priorizados en análisis y decisiones posteriores.

- **Atenciones por año:**

La visualización temporal evidencia volúmenes elevados y relativamente estables en los Años 2017, 2018 y 2019, seguidos de una caída significativa en 2020. Esta disminución constituye una ruptura estructural coherente con el contexto sanitario excepcional de ese año, lo que refuerza la necesidad de considerar factores exógenos en la interpretación de los datos y en el entrenamiento de modelos predictivos.

- **Atenciones por grupo de edad:**

La distribución por grupo etario revela una alta concentración en los rangos 00–04 Años y 30–59 Años, seguidos por la población joven adulta. Este patrón pone en evidencia la doble orientación del sistema hacia la atención materno-infantil y hacia la población adulta económicamente activa, con implicancias directas en políticas de prevención, planificación de recursos y sostenibilidad del sistema.

En conjunto, el dashboard descriptivo permite identificar de forma clara patrones estructurales, concentraciones críticas y variaciones temporales relevantes, estableciendo una línea base sólida y validada para el desarrollo de las fases predictiva y prescriptiva del modelo analítico avanzado.

Analítica Descriptiva del Sistema de Atenciones de Salud (SIS)

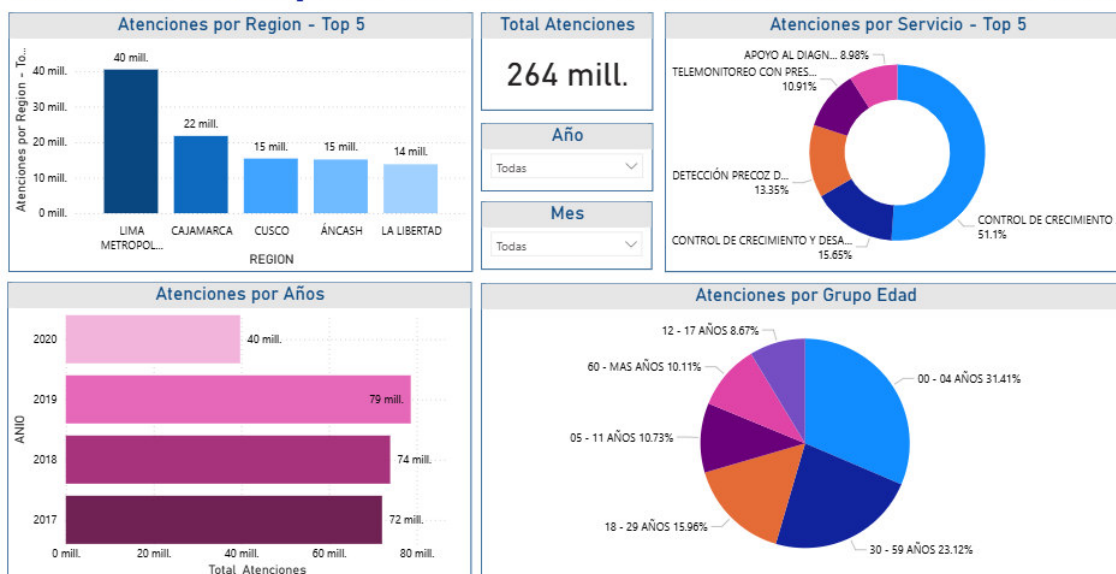


Figura 9 Dashboard Analítica Descriptiva

En la Figura 10 del Dashboard Predictivo muestra el comportamiento comparativo entre los valores reales observados y los valores predichos por el modelo para el sistema de atenciones de salud del SIS, utilizando información histórica agregada. Los resultados corresponden al año 2020, el cual fue reservado como conjunto de prueba (test set), mientras que los Años 2017, 2018 y 2019 fueron utilizados para el entrenamiento del modelo. Por tanto, los meses visualizados (marzo a diciembre) pertenecen exclusivamente al año 2020 y reflejan un escenario real de validación temporal, no una

mezcla de periodos.

En el panel superior izquierdo se presentan los principales indicadores de desempeño del modelo predictivo. El valor del MAE (109.43) indica el error promedio absoluto entre los valores reales y los predichos, mientras que el RMSE (697.66) evidencia la presencia de errores de mayor magnitud en ciertos escenarios específicos. La distribución de los registros entre entrenamiento (1,659,242) y prueba (459,346) confirma que el modelo fue evaluado sobre un volumen de datos significativo y temporalmente separado, reforzando la validez del análisis. Asimismo, se observa que el modelo se entrenó considerando los 40 servicios con mayor volumen de atenciones, lo que permitió reducir la complejidad sin perder representatividad del sistema.

El Cuadro central del dashboard muestra los casos con mayor error absoluto, destacándose principalmente escenarios de Lima Metropolitana, en el servicio de Consulta Externa, y concentrados en los grupos etarios de 30 a 59 Años y 60 Años a más. Estos registros corresponden a meses consecutivos del año 2020, lo cual explica la repetición de valores predichos constantes para un mismo escenario. Esta coherencia se debe a que el modelo aprende patrones históricos promedio para cada combinación de región, servicio y grupo etario, y no ajusta dinámicamente cada mes ante cambios abruptos en la demanda real.

En la gráfica inferior de líneas se presenta el Promedio de Predicho y Promedio de Real por Mes, donde se distinguen claramente dos comportamientos. La línea de color rojo representa el promedio de los valores predichos y se mantiene relativamente estable a lo largo de los meses, alrededor de los 96 mil a 116 mil atenciones, reflejando el patrón aprendido del histórico pre-2020. Por otro lado, la línea de color azul representa el promedio de los valores reales observados y muestra una caída pronunciada en el mes de abril, seguida de una recuperación progresiva hasta diciembre. Esta diferencia visual evidencia que el modelo no captura de forma precisa los quiebres estructurales ocurridos en 2020, sino que reproduce tendencias históricas consolidadas.

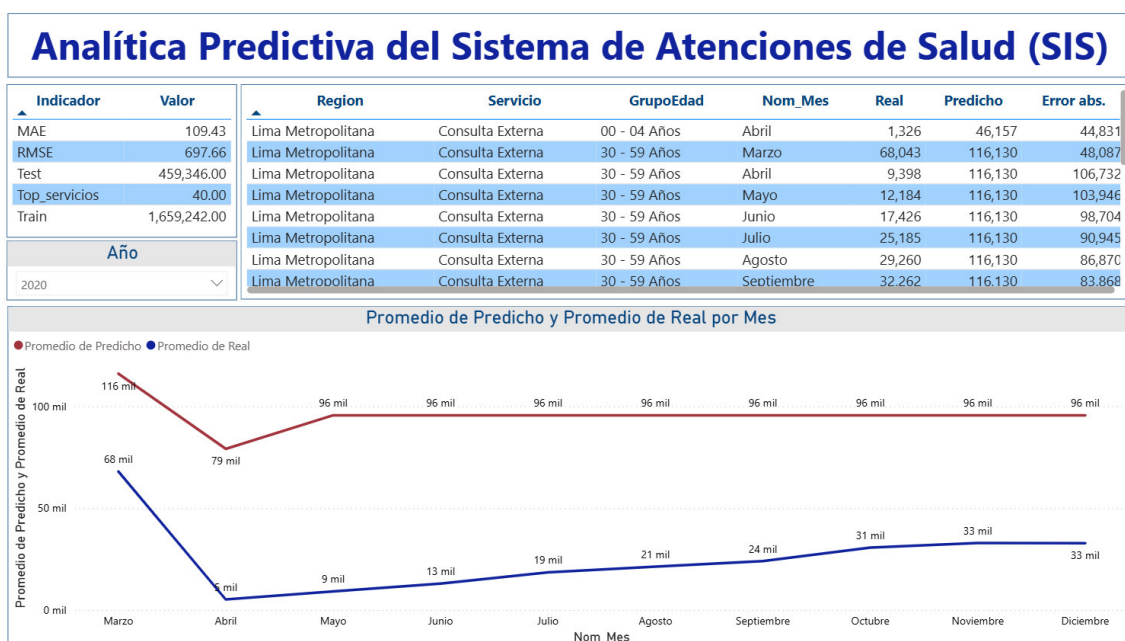


Figura 10 Dashboard Analítica Predictiva

En síntesis, el dashboard evidencia que el propósito del análisis predictivo no reside en estimar con exactitud milimétrica cada valor mensual, sino en anticipar configuraciones plausibles de demanda a partir de regularidades históricas. Las divergencias entre valores observados y proyectados no constituyen fallas conceptuales; operan, más bien, como señales analíticas que exigen lectura prudente y juicio experto. Tales insumos se integran de manera directa en la fase prescriptiva, donde los escenarios se jerarquizan y se formulan recomendaciones orientadas a respaldar la asignación estratégica de recursos sanitarios.

La Figura 11 expone el dashboard de analítica prescriptiva del Sistema de Atenciones de Salud (SIS), instancia en la que las proyecciones se traducen en lineamientos operativos. Estas directrices buscan fortalecer la toma de decisiones organizacionales en entornos de restricción presupuestaria, encauzando la asignación de esfuerzos bajo criterios explícitamente vinculados a la eficiencia.

Mientras la analítica descriptiva esclarece lo ya acontecido y la predictiva modeliza lo potencialmente venidero, la prescriptiva introduce un desplazamiento cualitativo: propone cursos de acción concretos y delimita modos de intervención. Desde esta perspectiva, la visualización del dashboard subraya aquellos escenarios con mayor

impacto esperado, habilitando la priorización de intervenciones allí donde el beneficio anticipado podría resultar más significativo.

- La gráfica de ranking de escenarios prioritarios (impacto esperado) muestra las combinaciones Región–Servicio con mayor volumen de atenciones proyectadas para el mes analizado (MonthIndex = 48). Este ranking permite identificar de forma directa dónde concentrar recursos, destacando escenarios como Consulta Externa en La Libertad, Cajamarca y Cusco, los cuales presentan la mayor presión futura estimada. Desde un enfoque prescriptivo, esta visualización guía la asignación priorizada de personal, horarios y capacidad operativa, evitando decisiones reactivas basadas únicamente en históricos agregados.
- La distribución de prioridades por región, representada mediante un gráfico circular, evidencia que la presión futura no es homogénea a nivel territorial. Regiones como Cajamarca y La Libertad concentran una proporción significativa de escenarios de alta prioridad, lo que refuerza la necesidad de decisiones diferenciadas por territorio. Esta visualización apoya la optimización de la toma de decisiones al permitir focalizar intervenciones regionales específicas, en lugar de aplicar políticas uniformes con menor efectividad.
- El gráfico de servicios que requieren acción prioritaria muestra que la mayor carga proyectada se concentra en servicios específicos, principalmente Consulta Externa y Control de Crecimiento y Desarrollo en Menores. Desde la analítica prescriptiva, este resultado no solo informa qué servicios son críticos, sino que indica cómo redistribuir recursos funcionales, como ampliación de turnos, refuerzo de atención primaria o redistribución de agendas médicas, maximizando el impacto de las decisiones adoptadas.
- El gráfico de población afectada (decisión focalizada) identifica los grupos etarios con mayor volumen de atenciones proyectadas, destacando principalmente a los grupos de 00–04 Años y 30–59 Años. Esta información permite optimizar la toma de decisiones clínicas y operativas al orientar acciones hacia poblaciones específicas, facilitando estrategias focalizadas como campañas preventivas, atención diferenciada

o priorización de programas materno-infantiles y de adultos en edad productiva.

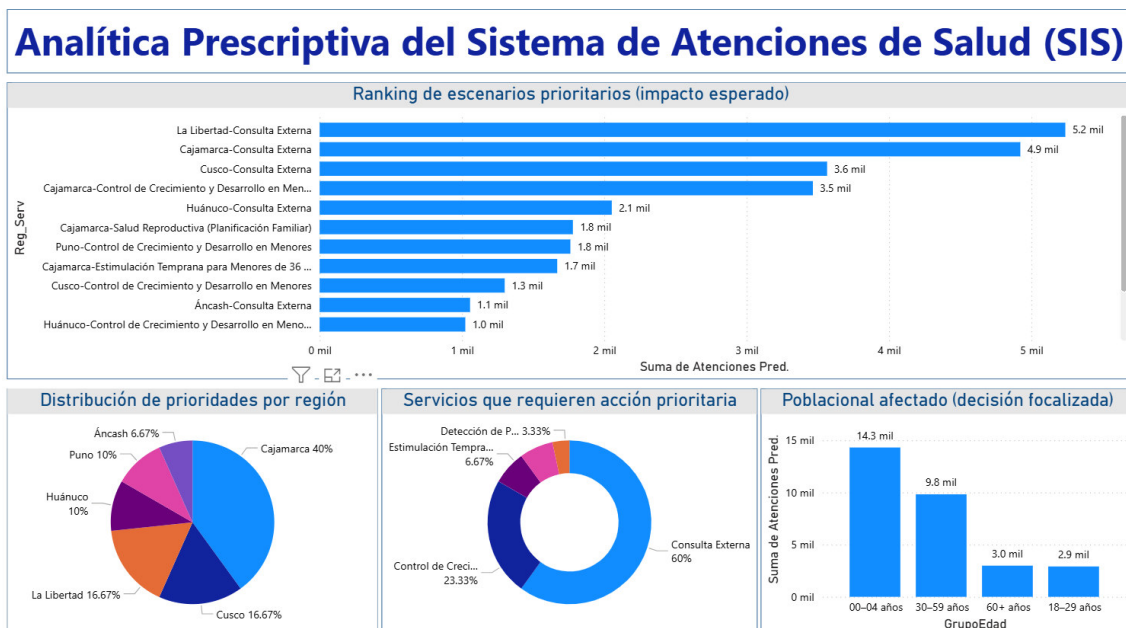


Figura 11 Dashboard Analítica Prescriptiva

En conjunto, el dashboard prescriptivo demuestra la eficacia de la analítica prescriptiva al traducir predicciones en decisiones accionables, proporcionando a los tomadores de decisión una guía clara sobre qué hacer, dónde hacerlo y sobre qué población actuar primero. Este enfoque contribuye directamente a optimizar la toma de decisiones (TD) al reducir la incertidumbre, priorizar intervenciones con mayor impacto esperado y maximizar el uso eficiente de los recursos del sistema de salud, evidenciando el valor práctico del modelo de analítica avanzada propuesto.

4.3.2 Comunicación (Fase DSR 6)

La fase de Comunicación de la metodología DSR se evidencia en la presentación y difusión de los resultados del modelo holístico desarrollado, como parte de una investigación doctoral. Los resultados del cuestionario aplicado a siete expertos indican que el modelo propuesto es conceptualmente coherente, metodológicamente sólido y pertinente para los objetivos planteados. La mayoría de los indicadores por pregunta se sitúan por encima de los umbrales de aceptación recomendados, con promedios superiores a 3.5 y niveles de acuerdo (valoraciones de 4 o 5) que superan el 60% en la mayoría de las dimensiones evaluadas.

Estos resultados respaldan la validez y aplicabilidad del modelo en contextos reales, constituyendo un aporte que se espera beneficie a la comunidad académica y profesional interesada en procesos de toma de decisiones basados en analítica avanzada. La publicación de este artículo científico representa un mecanismo formal de diseminación, cumpliendo así con los principios de la fase de comunicación de DSR.

CAPÍTULO 5: RESULTADOS Y DISCUSIÓN

5.1 Resultados

La presente sección tiene como propósito evaluar el desempeño del Modelo Holístico de Analítica Avanzada (MHAA) y demostrar, con base en evidencia empírica y en la validación experta, que el artefacto diseñado contribuye a la optimización del proceso de toma de decisiones organizacionales.

En consonancia con el enfoque de Design Science Research (DSR), no se recurre a una contrastación estadística de hipótesis en sentido positivista, sino a una evaluación sistemática del grado de cumplimiento de las proposiciones formuladas, sustentada en la triangulación metodológica entre el juicio experto y el caso aplicado.

5.1.1 Resultados del Juicio de Expertos

La validación mediante juicio de expertos se orientó a examinar la pertinencia conceptual, la utilidad práctica y la viabilidad operativa del Modelo Holístico de Analítica Avanzada. Participaron siete especialistas con trayectoria acreditada en analítica avanzada, arquitectura de datos y soporte decisonal en contextos organizacionales complejos. Los resultados obtenidos evidencian:

- **Pertinencia:** media superior a 4.0
- **Utilidad:** media superior a 4.2
- **Viabilidad:** media cercana a 4.0

Estos valores, en una escala Likert de 1 a 5, reflejan un alto nivel de consenso respecto a la coherencia estructural del modelo, su aplicabilidad práctica y su factibilidad de implementación en contextos organizacionales reales.

La Figura 12 presenta la representación gráfica tipo radar de los resultados del juicio de expertos. La visualización permite observar la consistencia entre las dimensiones evaluadas, destacando especialmente la utilidad del modelo, sin evidenciar desequilibrios significativos que comprometan su viabilidad o pertinencia conceptual.

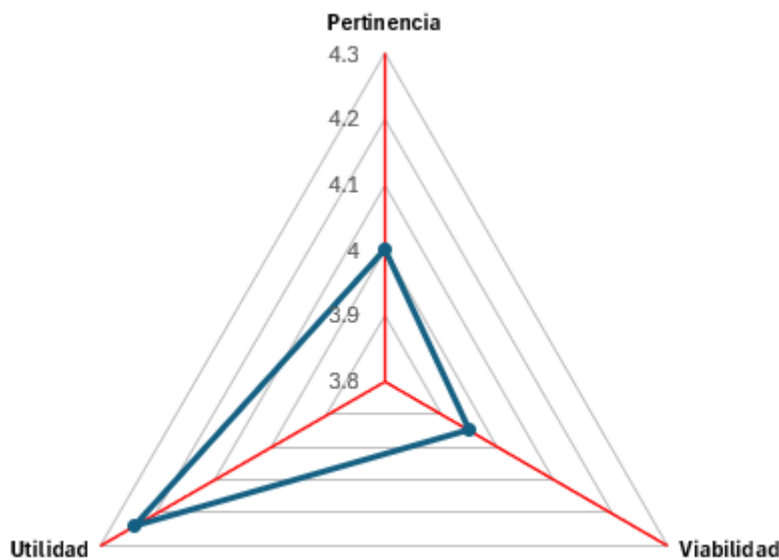


Figura 12 Resultados del juicio de expertos (pertinencia, viabilidad y utilidad)

Desde la perspectiva DSR, estos resultados validan el diseño conceptual y arquitectónico del modelo en términos de aceptación experta y adecuación estructural.

5.1.2 Resultados del Caso Aplicado

La ejecución del caso aplicado permitió evaluar el desempeño operativo del modelo en un entorno real de datos abiertos. A diferencia del juicio experto, que, valida la estructura, el caso aplicado valida la funcionalidad.

Los principales resultados fueron:

- Mejora en la calidad del dataset tras procesos de perfilado, limpieza y validación.
- Integración semántica y control de versiones en la fase de integración de datos.
- Implementación de análisis descriptivo, predictivo y prescriptivo dentro de un mismo flujo analítico.
- Generación de recomendaciones accionables.
- Desarrollo de dashboards estratégicos con indicadores clave de desempeño (KPIs).

Estos resultados evidencian que el modelo no solo fue conceptualizado, sino ejecutado de manera integral, manteniendo coherencia entre sus cinco capas

estructurales y demostrando capacidad analítica progresiva hasta llegar a la prescripción.

Desde el punto de vista de ingeniería de sistemas, esta ejecución constituye una validación funcional del artefacto.

5.1.3 Evaluación del Cumplimiento de las Propositiones

En coherencia con el enfoque DSR, la evaluación se realizó mediante verificación del cumplimiento de las proposiciones formuladas, como se muestra en la Tabla 32.

Tabla 32

Evaluación del cumplimiento de las proposiciones del MHAA

Proposición	Enunciado	Evidencia (Cap. 4)	Resultado
PR1 – Arquitectura Macro	La estructuración sistemática de las fases del modelo permite establecer una arquitectura analítica coherente y fundamentada teóricamente.	• Tabla 5: Estructura conceptual del modelo. • Integración sistemática de marcos teóricos (Simon, Davenport, DDDM, DAMA-DMBOK, Kimball, CRISP-DM, entre otros) • Validación experta con alto promedio en Pertinencia. • Ejecución completa de las cinco capas del modelo.	Confirmada
		• Integración progresiva desde contexto organizacional hasta visualización • Implementación operativa del flujo analítico end-to-end • Generación estructurada de insights. • Promedios superiores a 3.9 en pertinencia, viabilidad y utilidad	
PR2 – Diseño Funcional	La descomposición funcional de la arquitectura permite articular el modelo con el proceso de toma de decisiones organizacionales.	• Implementación de dashboards estratégicos • Generación de recomendaciones prescriptivas • Evidencia operativa en el caso aplicado.	Confirmada
PR3 – Validación Global	La validación mediante juicio de expertos y caso aplicado permite evidenciar la utilidad, viabilidad y eficacia del modelo en la optimización del proceso decisional.		

5.1.4 Triangulación de Evidencia

La optimización propuesta no se fundamenta en una única fuente de validación, sino en la convergencia de tres dimensiones de evidencia:

- Validación estructural (Juicio de Expertos): confirma la coherencia conceptual y la utilidad del diseño.
- Validación operativa (Caso Aplicado): demuestra la funcionalidad y eficacia técnica del artefacto.
- Alineación estructural (Tabla 11 de alineación con el modelo de TD): evidencia configuración compatible con el proceso decisional.

La triangulación de estas fuentes permite afirmar que el modelo presenta utilidad conceptual, eficacia operativa y una configuración alineada con los requisitos de optimización de la toma de decisiones organizacionales.

Prueba de Concepto (Proof of Concept – PoC)

La ejecución integral del caso aplicado constituye una prueba de concepto exitosa del Modelo Holístico de Analítica Avanzada. Se sintetiza la evidencia empírica en la Tabla 33.

Tabla 33***Matriz de Evidencia Empírica de la Prueba de Concepto***

Dimensión del Modelo	Evidencia Empírica	Resultado
Calidad de datos	Mejora del perfil y reducción de errores	Confirmada
Integración	Validación semántica y control de versiones	Confirmada
Analítica predictiva	Métricas de precisión satisfactorias	Confirmada
Analítica prescriptiva	Generación de recomendaciones accionables	Confirmada
Visualización	Implementación de dashboards con KPIs	Confirmada

5.1.5 Argumento de Optimización de la Toma de Decisiones

La optimización de la toma de decisiones se define bajo el enfoque de la Ingeniería de Sistemas como la configuración del artefacto para maximizar su utilidad y eficacia (Hevner et al., 2004), logrando satisfacer los requisitos de diseño bajo las restricciones del contexto organizacional (Simon, 1996).

En este sentido, el modelo:

- Reduce la incertidumbre inicial mediante depuración sistemática de datos.
- Integra múltiples capas analíticas en un solo flujo coherente.
- Genera prescripciones accionables basadas en evidencia.
- Permite retroalimentación continua.
- Articula gobernanza de datos con autonomía analítica.

La convergencia entre juicio de expertos y caso aplicado permite concluir que el artefacto analítico optimiza la toma de decisiones al transformar información dispersa en decisiones basadas en evidencia, bajo un marco estructurado y validado.

5.2 Discusión

La comparación del Modelo Holístico de Analítica Avanzada (MHAA) con la literatura revisada evidencia que la mayoría de los enfoques existentes abordan dimensiones parciales del proceso analítico. Modelos como CRISP-DM (Staneviciene et al., 2024; Xavier et al., 2025) y DataOps (Bahaa et al., 2023; AbouZaid et al., 2025) estructuran adecuadamente fases técnicas de minería o automatización, pero no integran explícitamente estándares de gobernanza, aseguramiento de calidad ni alineación formal con el proceso organizacional de toma de decisiones. De manera similar, arquitecturas clásicas de almacenamiento de datos como las propuestas por Kimball o Inmon se concentran en la estructuración dimensional y la integración de información, sin extenderse hacia niveles analíticos prescriptivos o autónomos ni vincularse metodológicamente con la decisión estratégica.

En el ámbito de la representación de la incertidumbre, el enfoque de Conjuntos Difusos Continuos Fraccionarios de Khan et al. (2025) aporta herramientas matemáticas robustas para la evaluación de alternativas bajo ambigüedad; sin embargo, su alcance se limita al cálculo metodológico y no contempla la gestión del ciclo de vida de los datos ni la integración con marcos de calidad y gobernanza. De igual modo, los modelos de decisión en contextos de crisis sanitaria propuestos por Romiti et al. (2025) enfatizan estilos de liderazgo y estructuras de gobernanza, pero no incorporan procesos formales de ingeniería de datos ni arquitecturas analíticas que soporten decisiones basadas en evidencia. Frente a estas aproximaciones, el MHAA integra calidad de datos (ISO 25012, ISO 8000), metadatos (ISO/IEC 11179), seguridad (ISO/IEC 27001), integración semántica y capacidades analíticas escalonadas —descriptivas, predictivas, prescriptivas y autónomas— dentro de un flujo coherente alineado con el proceso clásico de toma de decisiones.

Asimismo, marcos como la cadena de valor analítica de Davenport et al. (n.d.) y las propuestas que vinculan capacidades analíticas con la cadena de valor organizacional (Ghoshal et al., 2018; Bedeley et al., 2018) han permitido conceptualizar la progresión funcional del análisis. No obstante, estos enfoques permanecen en un plano clasificatorio

y no establecen una hoja de ruta operativa que articule estándares, limpieza de datos, integración, modelado y evaluación formal del artefacto. En contraste, el MHAA estructura dichas fases dentro de un marco metodológico DSR, permitiendo no solo la integración técnica sino también su validación empírica mediante juicio de expertos y caso aplicado.

En este sentido, mientras los modelos revisados contribuyen a mejorar aspectos específicos del proceso analítico o decisional, el MHAA optimiza estructuralmente la toma de decisiones organizacionales al configurar una arquitectura integral que reduce la fragmentación, articula gobernanza y analítica avanzada, y permite la generación de conocimiento accionable dentro de un flujo sistémico validado.

La Tabla 34 presenta una evaluación comparativa estructural de modelos de analítica respecto a la optimización de la Toma de Decisiones (TD). El criterio “Nivel Analítico Integral” considera los niveles D-P-Pr-A (Descriptivo, Predictivo, Prescriptivo y Autónomo). Asimismo, se evalúa la alineación explícita con el proceso de TD, la integración formal de estándares internacionales (ISO 25012, ISO 8000, ISO/IEC 11179, ISO/IEC 27001) y el carácter holístico del enfoque. El esquema semafórico sintetiza el grado de cumplimiento:  Verde = Cumple plenamente;  Amarillo = Cumple parcialmente;  Rojo = No cumple

Los resultados evidencian que los modelos tradicionales presentan coberturas parciales o no estructuradas en varios criterios, mientras que el MHAA integra de manera explícita los cuatro niveles analíticos, la alineación funcional con la TD y estándares de gobernanza, alcanzando un cumplimiento integral en los criterios evaluados.

Tabla 34

Evaluación comparativa estructural de modelos de analítica respecto a la optimización de la TD

Modelo	Nivel Analítico Integral (D-P-Pr-A)	Alineación explícita con TD	Integración de estándares (ISO / Gobernanza)	Enfoque Holístico	Evaluación Global
Kimball	● (Descriptivo predominante)	● No explícita	● No considerada	● Parcial (DW)	●
Inmon	● (Descriptivo predominante)	● No explícita	● No considerada	● Parcial (DW corporativo)	●
CRISP-DM	● (Predictivo predominante)	● Implícita	● No integrada formalmente	● Parcial (metodológico)	●
DDDM	● (Conceptual)	● Conceptual	● No estructurada	● No arquitectónico	●
MHAA (propuesto)	● Descriptivo + Predictivo + Prescriptivo + Autónomo	● Estructural y funcional	● ISO 25012, ISO 8000, ISO/IEC 11179, ISO/IEC 27001	● Integral	●

CONCLUSIONES

La investigación partió de un problema claramente delimitado: la fragmentación entre la analítica avanzada y el proceso de toma de decisiones organizacionales, evidenciada en entornos donde los datos se encuentran dispersos, con problemas de calidad y sin una arquitectura integrada que permita transformarlos en conocimiento accionable. Frente a esta brecha, el artefacto desarrollado —el Modelo Holístico de Analítica Avanzada (MHAA)— no se limitó a proponer una estructura conceptual, sino que articuló de manera operativa las fases de contexto organizacional, limpieza y calidad de datos, integración, analítica avanzada y visualización, alineándolas explícitamente con las fases de inteligencia, diseño y elección del modelo de Simon. Esta alineación estructural permitió transformar un escenario caracterizado por análisis fragmentados en un sistema integrado, capaz de generar una línea base descriptiva sólida, modelos predictivos con desempeño cuantificado y recomendaciones prescriptivas priorizadas. En este sentido, la pregunta de investigación —cómo el desarrollo de un modelo holístico de analítica avanzada optimiza la toma de decisiones organizacionales— encuentra respuesta en la evidencia empírica que demuestra que la arquitectura propuesta no solo organiza el flujo analítico, sino que lo conecta directamente con la lógica decisional organizativa.

El cumplimiento del objetivo general se sustenta en el logro progresivo de los objetivos específicos bajo la lógica iterativa de la Design Science Research. El objetivo específico 1 se concretó mediante la estructuración formal de la arquitectura del artefacto, cuya coherencia interna quedó representada en la configuración conceptual del modelo y su alineación con el proceso decisional (Tabla 11 en el documento). El objetivo específico 2 se materializó en la descomposición funcional del MHAA y en la especificación de sus componentes operativos, evidenciada en la ejecución real de cada fase sobre el dataset del SIS. Finalmente, el objetivo específico 3 se validó en dos niveles complementarios: primero, mediante juicio de expertos; segundo, a través de un caso aplicado con datos abiertos de gran escala. En el juicio de expertos, el modelo alcanzó una media de 4.00 en pertinencia, 3.95 en viabilidad y 4.24 en utilidad, con un 80.95% de valoraciones altas en esta última dimensión. Estos resultados no solo respaldan la proposición PR3, sino que evidencian la adecuación del artefacto al dominio organizacional y su potencial de

implementación. La evaluación por ítems mostró además tendencias favorables con medias predominantemente entre 3.5 y 4.5, confirmando una aceptación consistente del modelo

En el caso aplicado, la evaluación de utilidad y eficacia operativa se sustentó en resultados cuantificables. La fase descriptiva permitió establecer patrones temporales y segmentaciones etarias relevantes, consolidando una base analítica contextualizada. La fase predictiva reportó un MAE de 109.43 y un RMSE de 697.66 sobre una partición temporal (2017–2019 vs. 2020), con 1,659,242 registros de entrenamiento y 459,346 de prueba. Estos indicadores confirman la viabilidad técnica del enfoque predictivo bajo condiciones de alta heterogeneidad de datos. Asimismo, la identificación de errores concentrados en escenarios de alta demanda permitió interpretar críticamente los límites del modelo, reforzando su uso como soporte decisional y no como sustituto automático del juicio gerencial. La fase prescriptiva, sustentada en recomendaciones priorizadas (Tablas 29–31 en el documento), tradujo los resultados predictivos en escenarios de asignación estratégica de recursos, cerrando el ciclo decisional. En conjunto, la triangulación entre validación experta, métricas analíticas y evidencia operativa permitió sostener empíricamente el argumento de optimización de la toma de decisiones (sección 5.1.5 del documento).

Desde la perspectiva del conocimiento de diseño, la contribución doctoral trasciende la instancia concreta del artefacto. En primer lugar, el MHAA constituye una instanciación validada que demuestra cómo integrar gobernanza de datos, calidad, analítica avanzada y visualización bajo una arquitectura coherente con la lógica decisional organizativa. En segundo lugar, del proceso iterativo emergen principios de diseño generalizables: (i) la necesidad de una fase explícita de estructuración contextual previa a la modelización predictiva; (ii) la integración secuencial pero articulada de analítica descriptiva, predictiva y prescriptiva; (iii) la validación dual —experta y operativa— como mecanismo de legitimación del artefacto. En tercer lugar, el estudio aporta elementos que configuran una nascent design theory en el dominio de la analítica avanzada aplicada a la toma de decisiones: la optimización no depende únicamente de la sofisticación algorítmica, sino de la coherencia arquitectónica y de la alineación explícita entre fases analíticas y fases decisionales.

De este modo, la investigación amplía la knowledge base del campo al ofrecer una arquitectura validada empíricamente, criterios de evaluación operacionalizados y evidencia de desempeño en un entorno real de alta complejidad. La contribución no radica en afirmar que la analítica optimiza las decisiones —tesis ya explorada en la literatura—, sino en demostrar cómo debe diseñarse un artefacto integrado para que dicha optimización sea estructuralmente posible, evaluable y replicable bajo los principios de la Design Science Research.

RECOMENDACIONES

Las futuras implementaciones del Modelo Holístico de Analítica Avanzada deben preservar como condición estructural la coherencia entre gobernanza de datos, calidad, integración, capacidades analíticas y visualización, evitando configuraciones fragmentadas que disocien las fases analíticas del proceso organizacional de decisión. La optimización demostrada en esta investigación se sustenta en dicha integración arquitectónica y no en el uso aislado de herramientas técnicas.

Resulta pertinente extender la instanciación del modelo a entornos organizacionales caracterizados por mayor heterogeneidad estructural, alta velocidad de generación de datos o múltiples fuentes distribuidas, con el propósito de examinar la estabilidad de la arquitectura propuesta bajo condiciones de mayor complejidad operativa y presión analítica, fortaleciendo así su robustez intersectorial.

Asimismo, futuras investigaciones deberían incorporar mecanismos sistemáticos de seguimiento del desempeño del modelo mediante indicadores cuantificables que permitan evaluar periódicamente la precisión de los resultados analíticos y la consistencia de las recomendaciones generadas, contribuyendo a sostener el efecto optimizador en horizontes temporales extendidos.

Finalmente, el Modelo Holístico de Analítica Avanzada puede constituir la base para el desarrollo de infraestructuras tecnológicas escalables y desacopladas que automaticen integralmente el flujo analítico propuesto, desde la gestión de datos hasta la generación de recomendaciones accionables. Esta evolución permitiría reducir progresivamente la intervención humana en las fases operativas y posibilitar que el modelo opere en su nivel máximo de optimización estructural dentro de las organizaciones, sin comprometer coherencia, trazabilidad ni control decisional.

REFERENCIAS BIBLIOGRÁFICAS

- AbouZaid, A., Barclay, P. J., Chrysoulas, C., & Pitropakis, N. (2025). Building a modern data platform based on the data lakehouse architecture and cloud-native ecosystem. *Discover Applied Sciences*, 7, 166. <https://doi.org/10.1007/s42452-025-06545-w>
- Almeida, F., & Gama, J. (2017). *Knowledge discovery and data mining*. Springer.
- Alnajjar, O., Atencio, E., & Turmo, J. (2025). Framework for optimizing the construction process: The integration of lean construction, building information modeling (BIM), and emerging technologies. *Applied Sciences*, 15, 7253. <https://doi.org/10.3390/app15137253>
- Amann, W., & Gomez, P. (2014). *Developing executive leadership*. Routledge.
- Amir, A., Vinke, C., & van Heijningen, R. (2024). Horti-IoT: A new data framework for Dutch horticulture. <https://doi.org/10.22434/ifaamr1124>
- Aradiel Castañeda, H., Acosta De La Cruz, P. R., Mas Azahuanche, G. A., Geronimo Vasquez, A. H., & Flores Salinas, J. A. (2024). Generating dashboards with Python on the RPA process for optimizing the assignment of gardening tasks. In *Proceedings of the 4th LACCEI International Multiconference on Entrepreneurship, Innovation and Regional Development*. <https://doi.org/10.18687/LEIRD2024.1.1.709>
- Asikhia, O. U., & Mba, C. N. (2021). The influence of strategic decision making on organizational performance. *The International Journal of Business & Management*, 9(1). <https://doi.org/10.24940/theijbm/2021/v9/i1/BM2101-050>
- Bahaa, S., Ghalwash, A. Z., & Harb, H. (2023). DataOps lifecycle with a case study in healthcare. *International Journal of Advanced Computer Science and Applications*, 14. <https://doi.org/10.14569/IJACSA.2023.0140115>
- Baijens, J., Helms, R., & Bollen, L. (2022). Data analytics and SMEs: How maturity improves performance. In *Proceedings of the 2022 IEEE 24th Conference on Business Informatics (CBI)* (pp. 31–39). IEEE. <https://doi.org/10.1109/CBI54897.2022.00011>
- Bedeley, R. T., Ghoshal, T., Iyer, L. S., & Bhadury, J. (2018). Business analytics and organizational value chains: A relational mapping. *Journal of Computer Information Systems*, 58, 151–161. <https://doi.org/10.1080/08874417.2016.1220238>
- Belfo, F. P., & Andreica, A. B. (2018). A comprehensive methodology to implement business intelligence and analytics through knowledge discovery in databases. In A. Groza & R. Prasath (Eds.), *Mining Intelligence and Knowledge Exploration. MIKE 2018. Lecture Notes in Computer Science* (Vol. 11308). Springer, Cham. https://doi.org/10.1007/978-3-030-05918-7_10
- Benítez, Y. L. (2018). *Business intelligence: ADGG102PO*.

- Berge, G. T., Granmo, O. C., Tveit, T. O., Munkvold, B. E., Ruthjersen, A. L., & Sharma, J. (2023). Machine learning-driven clinical decision support system for concept-based searching: A field trial in a Norwegian hospital. *BMC Medical Informatics and Decision Making*, 23(1). <https://doi.org/10.1186/s12911-023-02101-x>
- Bhattacharjee, A. (2012). *Social science research: Principles, methods, and practices* (2nd ed.). University of South Florida.
- Borges, M., Bernardino, J., & Pedrosa, I. (2021). Data-driven decision making strategies applied to marketing. In *Proceedings of the 2021 16th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1–7). IEEE. <https://doi.org/10.23919/CISTI52073.2021.9476506>
- Bryman, A., & Bell, E. (2015). *Business research methods*. Oxford University Press.
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W.W. Norton & Company.
- Buitrago Pulido, R. D., Salinas, J., & Boude Figueroa, O. (2023). Validation of a model for the formalization of personal learning pathways through expert judgment. *Journal of Higher Education Theory and Practice*, 23(12). <https://doi.org/10.33423/jhetp.v23i12.6283>
- Caballero, I., Gualo, F., Rodríguez, M., & Piattini, M. (2022). BR4DQ: A methodology for grouping business rules for data quality evaluation. *Information Systems*, 109, 102058. <https://doi.org/10.1016/j.is.2022.102058>
- Candelo-Viáfara, J. M., & Gonzáles-Campo, C. H. (2022). Efecto de la incertidumbre en las organizaciones del mercado accionario: Una herramienta para la toma de decisiones y la inteligencia organizacional. *Estudios Gerenciales*, 57–68. <https://doi.org/10.18046/j.estger.2022.162.4689>
- Cardoso, E., & Su, X. (2022). Designing a business intelligence and analytics maturity model for higher education: A design science approach. *Applied Sciences (Basel, Switzerland)*, 12(9), 4625. <https://doi.org/10.3390/app12094625>
- Chacoma, A., & Billoni, O. V. (2025). Emergent complexity in the decision-making process of chess players. *Scientific Reports*, 15, 23234. <https://doi.org/10.1038/s41598-025-06335-9>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., y Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS Inc
- Chen, L., & Nath, R. (2018). Business analytics maturity of firms: An examination of the relationships between managerial perception of IT, business analytics maturity and success. *Information Systems Management*, 35(1), 62–77. <https://doi.org/10.1080/10580530.2017.1416948>
- Choudhary, S. (2024). Inventory optimization using an AI-powered holistic model for decision-

- support system. *International Journal for Research in Applied Science and Engineering Technology*, 12, 414–421. <https://doi.org/10.22214/ijraset.2024.58732>
- Clemen, R. T., & Reilly, T. (2001). *Making hard decisions with DecisionTools*. Duxbury Press.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.
- DAMA International. (2017). *DAMA-DMBOK: Data Management Body of Knowledge* (2nd ed.). Technics Publications.
- Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Davenport, T. H., & Harris, J. G. (2017). *Competing on analytics: The new science of winning* (Updated, with a new introduction). Harvard Business Review Press.
- Davenport, T. H., Harris, J., & Abney, D. (2017). *Competing on analytics: The new science of winning* (New ed.). Harvard Business Review Press.
- Davis, E. K., & Lee, S. H. (2022). Mejorar la colaboración en proyectos de análisis de datos utilizando matrices de coherencia. *Gestión de Sistemas de Información*, 39(3), 189–202.
- Demirdöğen, G., Diren, N. S., Aladağ, H., & Işık, Z. (2021). Lean based maturity framework integrating value, BIM and big data analytics: Evidence from AEC industry. *Sustainability*, 13(18), 10029. <https://doi.org/10.3390/su131810029>
- Drucker, P. F. (1967). *The effective executive: The definitive guide to getting the right things done*. HarperBusiness.
- Elgendy, N., Elragal, A., & Päivärinta, T. (2023). Evaluating collaborative rationality-based decisions: A literature review. *Procedia Computer Science*, 219, 647–657. <https://doi.org/10.1016/j.procs.2023.01.335>
- Evans, J. R. (2017). *Business analytics: Methods, models, and decisions*. Pearson.
- Few, S. (2018). *Data visualization: Communicating data and complex ideas effectively*. Analytics Press.
- Fivetran. (2022, May 11). Over 80 percent of companies rely on stale data for decision-making. *Business Wire*. <https://www.businesswire.com/news/home/20220511005403/en/Over-80-Percent-of-Companies-Rely-on-Stale-Data-for-Decision-Making>
- Flick, U. (2018). *An introduction to qualitative research*. SAGE Publications.
- Fowler, F. J. (2014). *Survey research methods* (5th ed.). Sage Publications.
- Frost, D., Gericke, O., Di Bari, R., Balangé, L., Zhang, L., Blagojevic, B., et al. (2022). Holistic quality model and assessment—Supporting decision-making towards sustainable construction using the design and production of graded concrete components as an example. *Sustainability*, 14, 11269. <https://doi.org/10.3390/su141811269>
- Gartner. (2016). *Data science and machine learning market study*.
- Gartner. (2022). *The Gartner top trends in data & analytics for 2022*.

- <https://www.gartner.com/en/webinar/430318/1035298>
- Ghoshal, T., Bedeley, R. T., Iyer, L. S., & Bhadury, J. (2018). Business analytics capabilities and use: A value chain perspective. In A. V. Deokar et al. (Eds.), *Analytics and data science: Advances in research and pedagogy* (pp. 41–54). Springer.
https://doi.org/10.1007/978-3-319-58097-5_5
- Gómez-Duque, L. Á., Daza-Torres, J. D., & Arias-Pérez, J. (2023). Inteligencia de negocios y agilidad organizacional: ¿Son relevantes la toma de decisiones racional e intuitiva? *Estudios Gerenciales*, 181–191. <https://doi.org/10.18046/j.estger.2023.167.5542>
- Grossman, R. L. (2018). A framework for evaluating the analytic maturity of an organization. *International Journal of Information Management*, 38(1), 45–51.
<https://doi.org/10.1016/j.ijinfomgt.2017.08.005>
- Gualo, F., Caballero, I., Rodríguez, M., & Piattini, M. (2023). A data quality model for master data repositories. *Informatica*, 34, 795–824. <https://doi.org/10.15388/23-INFOR534>
- Guillemot, J., Ortiz, A., Valdivieso, E., Muñoz, G., & Cisneros-Heredia, D. F. (2024). Approach for participatory Delphi method for structuring voices, population empowerment and research training. *MethodsX*, 13, 103019. <https://doi.org/10.1016/j.mex.2024.103019>
- Gupta, N., & Mangla, R. (2020). *Artificial intelligence basics: A self-teaching introduction*. Mercury Learning & Information.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. Pearson.
- Hallo, L., & Nguyen, T. (2022). Holistic view of intuition and analysis in leadership decision-making and problem-solving. *Administrative Sciences*, 12(1), 4.
<https://doi.org/10.3390/admsci12010004>
- Han, S. Y., Yoo, J., Michels, A., Kang, J.-Y., Wang, S., & Kim, J.-S. (2025). Vulnerable neighborhood explorer (VNE): An open-source visual analytics tool for exploring social vulnerability to disasters across different neighborhoods. *SoftwareX*, 31, 102233.
<https://doi.org/10.1016/j.softx.2025.102233>
- Harrati, N., Bouchrika, I., Tari, A., & Ladjailia, A. (2016). Exploring user satisfaction for e-learning systems via usage-based metrics and system usability scale analysis. *Computers in Human Behavior*, 61, 463–471. <https://doi.org/10.1016/j.chb.2016.03.051>
- Hegselmann, S., Storck, M., Gessner, S., Neuhaus, P., Varghese, J., Bruland, P., et al. (2021). Pragmatic MDR: A metadata repository with bottom-up standardization of medical metadata through reuse. *BMC Medical Informatics and Decision Making*, 21, 160.
<https://doi.org/10.1186/s12911-021-01524-8>
- Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian Journal of Information Systems*, 19(2), 87–92.

- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
- Hevner, A., & Chatterjee, S. (2010). Design science research in information systems. In A. Hevner & S. Chatterjee (Eds.), *Design research in information systems: Theory and practice* (pp. 9–22). Springer. https://doi.org/10.1007/978-1-4419-5653-8_2
- Hevner, A., & Chatterjee, S. (2012). *Design research in information systems: Theory and practice*. Springer.
- Hindle, G. A., & Vidgen, R. (2018). Developing a business analytics methodology: A case study in the foodbank sector. *European Journal of Operational Research*, 268(3), 836–851. <https://doi.org/10.1016/j.ejor.2017.06.031>
- Hu, Z., Wang, A., Duan, Y., Zhou, J., Hu, W., & Wu, S. (2024). Toward better semantic interoperability of data element repositories in medicine: Analysis study. *JMIR Medical Informatics*, 12, e60293. <https://doi.org/10.2196/60293>
- Hung, S.-Y., Chen, C., Choi, H. S., & Ractham, P. (2021). A holistic framework to examine the impact of user, organizational and data factors on the use of big data analytics systems. <https://doi.org/10.47989/irpaper915>
- IBM. (2013). The four V's of big data. IBM Big Data & Analytics Hub.
- Jeyaprabha, B., Kumar, S. R., Bolla, R. L., Bhatt, A. S., Sera, R. J., & Arora, K. (2025). Data-driven decision making in management: Leveraging big data analytics for strategic planning. In *Proceedings of the 2025 First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies* (pp. 1000–1003). IEEE. <https://doi.org/10.1109/CE2CT64011.2025.10939548>
- Jones, G. R., & George, J. M. (2010). *Administración contemporánea* (6a. ed.). McGraw-Hill Interamericana de España S. L.
- Kahneman, D., & Tversky, A. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Kang, K. (2020). Reformulating ‘holism’ in hydropower decision-making. *Systems Research and Behavioral Science*, 37, 360–368. <https://doi.org/10.1002/sres.2635>
- Khan, A., Khan, S., Rahimzai, A. A., & Abdullah, S. (2025). Extended version of decision making model for industrial robot selection via fractional continuous fuzzy information. *Journal of Big Data*, 12, 158. <https://doi.org/10.1186/s40537-025-01190-x>
- Kimball, R., Reeves, L., Ross, M., & Thornthwaite, W. (2008). *The data warehouse lifecycle toolkit* (2nd ed.). Wiley.
- Kolar, E., Tušak, M., Robnik, P., & Biloslavo, R. (2024). Strategic decisions: How good does it fit the decision-making process into the strategic management process? *Journal of Engineering Management and Competitiveness*, 14, 99–111. <https://doi.org/10.5937/JEMC2402099K>

- Komóczy, M., & Kósa, K. (2025). Storytelling as narrative health promotion in community psychiatry: A quasi-experimental study. *BMC Psychiatry*, 25, 376.
<https://doi.org/10.1186/s12888-025-06816-1>
- Korsten, G., Aysolmaz, B., Turetken, O., Edel, D., & Ozkan, B. (2022). ADA-CMM: A capability maturity model for advanced data analytics. *Proceedings of the Annual Hawaii International Conference on System Sciences*.
<https://doi.org/10.24251/HICSS.2022.032>
- Kuchta, D., Canonico, P., Capone, V., & Capaldo, G. (2023). Uncertainty in the planning phase of public projects—Its scope, consequences, and possible remedies. *Administrative Sciences*, 13(6), 145. <https://doi.org/10.3390/admsci13060145>
- Kulkarni, A. J., Siarry, P., Singh, P. K., Abraham, A., Zhang, M., Zomaya, A., & Baki, F. (2019). *Big data analytics in healthcare*. Springer Nature.
- Kvale, S., & Brinkmann, S. (2015). *InterViews: The craft of qualitative research interviewing* (3rd ed.). Sage Publications.
- Lemsa, S. (2021). Challenges of advanced analytics maturity model development. *Environment Technology Resources Proceedings of the International Scientific and Practical Conference*, 2, 88–92. <https://doi.org/10.17770/etr2021vol2.6621>
- Liberatore, M. J., & Wagner, W. P. (2022). Simon’s decision phases and user performance: An experimental study. *Journal of Computer Information Systems*, 62(4), 667–679.
<https://doi.org/10.1080/08874417.2021.1878476>
- Lin, J., Harada, K., & Goto, H. (2023). Designing and implementing organizational learning analytics following a holistic planner tool. <https://doi.org/10.2139/ssrn.4655662>
- Liu, S., Liu, O., & Chen, J. (2023). A review on business analytics: Definitions, techniques, applications and challenges. *Mathematics*, 11(4), 899.
<https://doi.org/10.3390/math11040899>
- Liu, X., Lu, W., & Zhang, J. (2020). Data governance framework for big data analytics in smart cities. *Journal of Urban Informatics*, 12(3), 245–259.
- Lock, D. (2020). *Project management*. Routledge.
- Malatji, M. (2023). Management of enterprise cyber security: A review of ISO/IEC 27001:2022. In *Proceedings of the International Conference on Cyber Management and Engineering* (pp. 117–122). IEEE.
<https://doi.org/10.1109/CyMaEn57228.2023.10051114>
- Manoharan, S. G. S., Subramaniam, R., & Mohapatra, S. (2023). The decision-making process at organizations. In *Enabling Strategic Decision-Making in Organizations Through Dataplex* (pp. 1–25). Emerald Publishing Limited.
- March, S. T., & Smith, G. F. (1995). *Design and natural science research on information*

- technology. *Decision Support Systems*, 15(4), 251–266.
- Mikša, M., & Sikirica, N. (2024). The evolution and applications of the technology acceptance model. *Management, Enterprise and Benchmarking in the 21st Century*, 4, 154–157. [https://doi.org/10.15341/mce\(2470-4180\)/05-06.04.2024/005](https://doi.org/10.15341/mce(2470-4180)/05-06.04.2024/005)
- Miller, R., Chan, S. H. M., Whelan, H., & Gregório, J. (2025). A comparison of data quality frameworks: A review. *Big Data and Cognitive Computing*, 9, 93. <https://doi.org/10.3390/bdcc9040093>
- Ministerio de Salud del Perú. (2017). Datos de atenciones realizadas a los asegurados – Seguro Integral de Salud (SIS). Plataforma Nacional de Datos Abiertos. <https://www.datosabiertos.gob.pe/dataset/datos-de-atenciones-realizadas-los-asegurados-sis>
- Morales-Arevalo, J. C., & Rodríguez, C. (2025). A systematic review of the benefits and challenges of data analytics in organizational decision making. *International Journal of Advanced Computer Science and Applications*, 16. <https://doi.org/10.14569/IJACSA.2025.0160221>
- Morales-Arevalo, J. C., Aquis-Gonzales, E. M., Carpio-Ore, W. Y., Sáenz, E. V. M., Mazzarri-Rodriguez, C. J., Remotti-Becerra, E. E., et al. (2025). Data mart design to increase transactional flow of debit and credit card in Peruvian bodegas. *International Journal of Advanced Computer Science and Applications*, 16. <https://doi.org/10.14569/IJACSA.2025.0160216>
- Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, 1(1), 48–76.
- Muñoz, M. A., Pineda, S., & Morales, J. M. (2022). A bilevel framework for decision-making under uncertainty with contextual information. *Omega*, 108, 102575. <https://doi.org/10.1016/j.omega.2021.102575>
- Nunes, F., Alexandre, E., & Gaspar, P. D. (2024). Implementing key performance indicators and designing dashboard solutions in an automotive components company: A case study. *Administrative Sciences*, 14, 175. <https://doi.org/10.3390/admsci14080175>
- Nutt, P. (2002). *Why decisions fail: Avoiding the blunders and traps that lead to debacles*. Berrett-Koehler Publishers.
- Organización Internacional de Normalización. (2019). ISO 31010. Gestión del riesgo—Técnicas de apreciación (Vol. 31010). ISO.
- Patel, A. D., & Kumar, S. (2023). Promover la transparencia en el análisis de datos a través de matrices de consistencia. *Ética en la Ciencia de Datos*, 7(2), 55–70.
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management*

- Information Systems, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Peñaloza, M. (2010). Teoría de las decisiones. *Perspectivas*, 25, 227–240.
- Pochiraju, B., & Seshadri, S. (Eds.). (2019). *Essentials of business analytics: An introduction to the methodology and its applications*. Springer International Publishing.
- Ponniah, P. (2010). *Data warehousing fundamentals for IT professionals*. Wiley.
- Prajapati, V. (2013). *Big data analytics with R and Hadoop*. Packt Publishing.
- Pratt, M. C., Jones, R. K., & Smith, L. (2020). *Data architecture and feasibility in advanced analytics implementation*. Academic Press.
- Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1), 51–59. <https://doi.org/10.1089/big.2013.1508>
- Quitizaca-Farfan, L., Vásquez-Bazaran, C., & Uriguen-Aguirre, P. (2022). Influencia del entorno externo e interno en la toma de decisiones gerenciales. *593 Digital Publisher CEIT*, 7(2), 339–354. <https://doi.org/10.33386/593dp.2022.2.1046>
- Rafique, M., Selamat, A., & Maqbool, U. (2019). Preprocessing framework for data quality improvement in big data analytics. *Journal of Big Data*, 6(7), 1–18.
- Reis, J., & Housley, M. (2022). *Fundamentals of data engineering: Plan and build robust data systems*. O'Reilly Media.
- Romiti, A., Del Vecchio, M., Cavicchi, C., & Vagnoni, E. (2025). Healthcare organizations in crisis context: Decision-making models and roles of CEOs. *BMC Health Services Research*, 25, 273. <https://doi.org/10.1186/s12913-025-12420-6>
- Ross, J. W., Sebastian, I. M., Beath, C., Mocker, M., Moloney, K. G., & Fonstad, N. O. (2019). *Designed for digital: How to architect your business for sustained success*. MIT Press.
- Saravanabhavan, H., Riaz, S., & Das, S. (2022). Understanding business analytics maturity drivers for Indian companies. In *2022 International Conference on Data Analytics for Business and Industry (ICDABI)*. <https://doi.org/10.1109/ICDABI56818.2022.10041566>
- Sateesh Reddy, C., & Nimo, J. R. A. (2023). Analyzing the impact of analytics on organizational performance. *International Journal of All Research Education and Scientific Methods*, 12, 3383–3390. <https://doi.org/10.56025/IJARESM.2023.1201243383>
- Saunders, M., Lewis, P., & Thornhill, A. (2016). *Research methods for business students* (7th ed.). Pearson.
- Sharda, R., Delen, D., & Turban, E. (2020). *Analytics, data science, & artificial intelligence: Systems for decision support* (11^a ed.). Pearson.
- Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl, K. C. (2017). *Data mining for business analytics: Concepts, techniques, and applications in R*. John Wiley & Sons.

- Simon, H. A. (1960). *The new science of management decision*. Harper & Brothers.
<https://doi.org/10.1037/13978-000>
- Simon, H. A. (1996). *The sciences of the artificial* (3.^a ed.). MIT Press.
- Song, Y., Li, W., Wang, Z., & Hu, Y. (2025). Visual storytelling through the void: A quantitative analysis of empty shot distribution across film genres. *Humanities and Social Sciences Communications*, 12, 232. <https://doi.org/10.1057/s41599-025-04527-2>
- Stander, H.-M., Cohen, B., Harrison, S. T. L., & Broadhurst, J. L. (2023). Validity of using expert judgements to inform multiple criteria decision analysis: Selecting technologies for sulfide-enriched fine coal waste reuse. *Journal of Cleaner Production*, 415, 137671. <https://doi.org/10.1016/j.jclepro.2023.137671>
- Staneviciene, E., Gudoniene, D., Punys, V., & Kukstys, A. (2024). A case study on the data mining-based prediction of students' performance for effective and sustainable e-learning. *Sustainability*, 16, 10442. <https://doi.org/10.3390/su162310442>
- Stubbs, E. (2014). *The value of business analytics: Identifying the path to profitability*. John Wiley & Sons.
- Szukits, Á. (2022). The illusion of data-driven decision making: The mediating effect of digital orientation and controllers' added value in explaining organizational implications of advanced analytics. *Journal of Management Control*, 33, 403–446. <https://doi.org/10.1007/s00187-022-00343-w>
- Tashakkori, A., & Teddlie, C. (2010). *Sage handbook of mixed methods in social & behavioral research*. Sage Publications.
- Tidd, J., & Bessant, J. (2013). *Managing innovation: Integrating technological, market and organizational change*. Wiley.
- Tuanama-Tuanama, K. L., De-La-Matta-Riva, J. P., Moreto-Correa, N., & Cruz-Tarrillo, J. J. (2022). La toma de decisiones (TD) en relación con la eficacia organizativa en empresas: Un artículo de revisión. *UNAAACIENCIA-PERÚ*, 1(1), e13. <https://doi.org/10.56926/unaaaciencia.v1i1.13>
- Utomo, W. H., Rosalina, R., & Afriyadi, A. (2024). Privacy-preserving data mining optimization for big data analytics using deep reinforcement learning. *Indonesian Journal of Electrical Engineering and Computer Science*, 36, 1929–1937. <https://doi.org/10.11591/ijeecs.v36.i3.pp1929-1937>
- Van Aken, J. E. (2004). Management research based on the design science paradigm: The quest for field-tested and grounded technological rules. *Journal of Management Studies*, 41(2), 219–246. <https://doi.org/10.1111/j.1467-6486.2004.00430.x>
- Vásquez, J., Aguirre, S., Puertas, E., Bruno, G., Priarone, P. C., & Settineri, L. (2021). A sustainability maturity model for micro, small and medium-sized enterprises (MSMEs)

- based on a data analytics evaluation approach. *Journal of Cleaner Production*, 311(127692), 127692. <https://doi.org/10.1016/j.jclepro.2021.127692>
- Von Alan, R. H., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105. <https://doi.org/10.2307/25148625>
- Wilder, C. (2015). The analytics mandate. *MIT Sloan Management Review*.
- Xavier, M. K., Leal, G. C. L., Guerino, G. C., Coleti, T. A., & Balancieri, R. (2025). Automatic complaints classification in e-commerce: A case study using CRISP-DM. *Journal on Interactive Systems*, 16, 256–266. <https://doi.org/10.5753/jis.2025.4661>
- Yan, L., Martinez-Maldonado, R., Jin, Y., Echeverria, V., Milesi, M., Fan, J., et al. (2025). The effects of generative AI agents and scaffolding on enhancing students' comprehension of visual learning analytics. *Computers & Education*, 234, 105322. <https://doi.org/10.1016/j.compedu.2025.105322>
- Yessad, L., & Labiod, A. (2016). Comparative study of data warehouses modeling approaches: Inmon, Kimball and Data Vault. In *Proceedings of the International Conference on System Reliability and Science* (pp. 95–99). IEEE. <https://doi.org/10.1109/ICSRS.2016.7815845>
- Yin, R. K. (2017). *Case study research and applications: Design and methods*. Sage Publications.
- Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2013). *Business research methods*.

ANEXO A: Código en Python Fase Estructura de Contexto Organizacional

```
# =====
# BLOQUE 1: Contexto Organizacional
# Planificación/objetivos: usar SIS 2017-2020 para construir una línea
# base de atenciones,
# definir KPIs y preparar una base integrada que permita análisis
# descriptivo + predicción + prescripción +
# Visualización y recomendaciones simples (eficacia del modelo).
# =====

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from pathlib import Path

# 1) Definir archivos (nombres tal cual se descargaron del repositorio)
files = {
    2018: "DataSet/OPENDATA_DS_01_2018_ATENCIONES.csv",
    2017: "DataSet/OPENDATA_DS_01_2017_ATENCIONES.csv",
    2019: "DataSet/OPENDATA_DS_01_2019_ATENCIONES.csv",
    2020: "DataSet/OPENDATA_DS_01_2020_ATENCIONES.csv",
}

# 2) Cargar y unificar (manteniendo trazabilidad por año/archivo)
dfs = []
for year, fname in files.items():
    path = Path(fname)
    if not path.exists():
        raise FileNotFoundError(f"No se encontró el archivo:
{path.resolve()}")

    # encoding: en open data peruano a veces es latin-1; si falla,
    # cambia a encoding="latin-1"
```

```

df = pd.read_csv(path, encoding="utf-8", low_memory=False)
df["ANIO_ARCHIVO"] = year # trazabilidad de origen
dfs.append(df)

data_raw = pd.concat(dfs, ignore_index=True)

# 3) Inspección inicial (contexto del dataset)
print("✅ Dataset integrado SIS (2017-2020)")
print("Filas/Columnas:", data_raw.shape)
print("\nColumnas (primeras 40):")
print(list(data_raw.columns)[:40])

# 4) Validación rápida de Años (si existe la columna AÑO)
if "AÑO" in data_raw.columns:
    print("\nAnios detectados en columna AÑO:",
sorted(data_raw["AÑO"].dropna().unique()))
else:
    print("\n⚠ No existe columna 'AÑO'. Se usará 'ANIO_ARCHIVO' como
año de referencia.")
    print("Anios por ANIO_ARCHIVO:",
sorted(data_raw["ANIO_ARCHIVO"].dropna().unique()))

# 5) Vista previa
display(data_raw.head(5))

```

ANEXO B: Código en Python Fase Limpieza y Calidad de Datos

```

# =====
# BLOQUE 2: Limpieza y Calidad de Datos
# -----
# Elementos del bloque (modelo):
# - Perfilado de datos: entender estructura, nulos, tipos, rangos, duplicados.

```

```

# - Evaluación de calidad: completitud, consistencia, validez y outliers.
# - Reglas de calidad y validación: reglas explícitas (dominios/catálogos).
# - Limpieza de datos: tratamiento de nulos, normalización de textos/códigos.
# - Normalización y estandarización: formatos homogéneos (tiempo, texto, IDs).
# - Consolidación de duplicados: agregación/control a nivel analítico.
# - Metadatos de calidad: reporte de KPIs de calidad (antes/después).
# - Monitoreo continuo: funciones reutilizables para re-ejecutar controles.
# =====

import pandas as pd
import numpy as np

# (Opcional pero recomendado para ver todo)
pd.set_option("display.max_columns", 200)
pd.set_option("display.width", 200)

# -----
# 0) Supuesto: df ya existe (salida del Bloque 1)
# df = dataset integrado 2017-2020
# -----

# -----
# 1) Selección de variables clave (alineadas a TD)
# Nota: mantenemos variables con valor analítico real.
# -----

key_columns = [
    "AÑO", "MES", "ANIO_ARCHIVO",
    "REGION", "PROVINCIA", "DISTRITO", "UBIGEO_DISTRITO",
    "SEXO", "GRUPO_EDAD",
    "COD_SERVICIO", "DESC_SERVICIO",
    "ATENCIONES",
    "IPRESS", "NIVEL_EESS"
]

# Verificar columnas existentes y filtrar solo las que estén presentes
existing_cols = [c for c in key_columns if c in df.columns]
missing_cols = [c for c in key_columns if c not in df.columns]

print("Columnas usadas (presentes):", existing_cols)
if missing_cols:

```

```

print("Columnas esperadas pero NO encontradas (revisar nombres):", missing_cols)

df2 = df[existing_cols].copy()

# -----
# 2) Perfilado inicial (calidad base)
# -----

def quality_profile(data: pd.DataFrame) -> pd.DataFrame:
    """Perfil simple de calidad por columna."""
    n = len(data)
    prof = []
    for col in data.columns:
        s = data[col]
        prof.append({
            "columna": col,
            "tipo": str(s.dtype),
            "n_filas": n,
            "nulos": int(s.isna().sum()),
            "%_nulos": float(s.isna().mean() * 100),
            "unicos": int(s.nunique(dropna=True)),
            "%_unicos": float((s.nunique(dropna=True) / n) * 100) if n else 0.0
        })
    return pd.DataFrame(prof).sort_values(["%_nulos", "unicos"], ascending=[False, True])

profile_before = quality_profile(df2)
print("\n=== Perfil de calidad (ANTES) - Top 15 por % nulos ===")
print(profile_before.head(15))

# -----
# 3) Estandarización de tipos y formatos
# -----

# Asegurar numéricos
for col in ["AÑO", "MES", "ANIO_ARCHIVO", "ATENCIONES"]:
    if col in df2.columns:
        df2[col] = pd.to_numeric(df2[col], errors="coerce")

# Estandarizar texto (strip + mayúsculas) en campos categóricos clave
text_cols = ["REGION", "PROVINCIA", "DISTRITO", "DESC_SERVICIO", "SEXO",
             "GRUPO_EDAD", "IPRESS", "NIVEL_EESS"]
for col in text_cols:

```

```

if col in df2.columns and df2[col].dtype == "object":
    df2[col] = df2[col].astype(str).str.strip().str.upper()

# UBIGEO como string "limpio" (para no perder ceros a la izquierda si los hubiera)
if "UBIGEO_DISTRITO" in df2.columns:
    df2["UBIGEO_DISTRITO"] = df2["UBIGEO_DISTRITO"].astype(str).str.strip()

# -----
# 4) Reglas de calidad (validez y consistencia básica)
# -----

quality_issues = {}

# Regla 1: AÑO razonable (2017-2020 para el caso aplicado)
if "AÑO" in df2.columns:
    invalid_year = df2[~df2["AÑO"].between(2017, 2020, inclusive="both") & df2["AÑO"].notna()]
    quality_issues["invalid_year_count"] = len(invalid_year)

# Regla 2: MES válido (1-12)
if "MES" in df2.columns:
    invalid_month = df2[~df2["MES"].between(1, 12, inclusive="both") & df2["MES"].notna()]
    quality_issues["invalid_month_count"] = len(invalid_month)

# Regla 3: ATENCIONES no negativas y no nulas
if "ATENCIONES" in df2.columns:
    neg_att = df2[df2["ATENCIONES"].notna() & (df2["ATENCIONES"] < 0)]
    quality_issues["negative_atenciones_count"] = len(neg_att)

# Regla 4: SEXO dominio esperado (puede variar: M/F, H/M, etc.)
if "SEXO" in df2.columns:
    # Dominio flexible: detectamos principales valores y marcamos raros
    top_sexo = df2["SEXO"].value_counts(dropna=True).head(10)
    quality_issues["top_sexo_values"] = top_sexo.to_dict()

# Regla 5: Claves de servicio (si existe COD_SERVICIO, que no sea nulo cuando DESC_SERVICIO
# no lo es)
if "COD_SERVICIO" in df2.columns and "DESC_SERVICIO" in df2.columns:
    inconsistent_service = df2[df2["DESC_SERVICIO"].notna() & df2["COD_SERVICIO"].isna()]
    quality_issues["service_inconsistency_count"] = len(inconsistent_service)

print("\n=== Chequeos de calidad (ANTES de limpiar nulos/outliers) ===")

```

```

for k, v in quality_issues.items():
    print(f"{k}: {v}")

# -----
# 5) Limpieza: tratamiento de nulos críticos
# (no inventamos datos; filtramos lo indispensable para análisis)
# -----

critical_cols = [c for c in ["AÑO", "MES", "REGION", "ATENCIONES", "COD_SERVICIO"] if c in
df2.columns]
df_clean = df2.dropna(subset=critical_cols).copy()

# ATENCIONES: si viniera como float por NaN, lo pasamos a int (seguro)
if "ATENCIONES" in df_clean.columns:
    df_clean = df_clean[df_clean["ATENCIONES"] >= 0] # elimina negativos
    df_clean["ATENCIONES"] = df_clean["ATENCIONES"].astype("int64", errors="ignore")

# Filtrar meses/Años inválidos
if "AÑO" in df_clean.columns:
    df_clean = df_clean[df_clean["AÑO"].between(2017, 2020, inclusive="both")]
if "MES" in df_clean.columns:
    df_clean = df_clean[df_clean["MES"].between(1, 12, inclusive="both")]

# -----
# 6) Consolidación de duplicados (a nivel analítico)
# Nota: el SIS puede tener múltiples filas por combinaciones.
# Para TD, normalmente agregamos por (tiempo, región, servicio, sexo, edad).
# -----

group_dims = [c for c in ["AÑO", "MES", "REGION", "PROVINCIA", "COD_SERVICIO",
"DESC_SERVICIO", "SEXO", "GRUPO_EDAD"] if c in df_clean.columns]

df_agg = (
    df_clean
    .groupby(group_dims, dropna=False, as_index=False)["ATENCIONES"]
    .sum()
)

# -----
# 7) Metadatos de calidad (KPIs antes/después)
# -----

def quality_kpis(before: pd.DataFrame, after: pd.DataFrame) -> pd.DataFrame:

```

```

return pd.DataFrame({
    "dataset": ["antes", "despues_filtrado", "despues_agregado"],
    "filas": [len(before), len(after), np.nan],
    "columnas": [before.shape[1], after.shape[1], np.nan]
})

kpis = pd.DataFrame({
    "etapa": ["original (df2)", "limpio (df_clean)", "agregado (df_agg)"],
    "filas": [df2.shape[0], df_clean.shape[0], df_agg.shape[0]],
    "columnas": [df2.shape[1], df_clean.shape[1], df_agg.shape[1]],
})

print("\n=== KPIs de calidad y transformación ===")
print(kpis)

profile_after = quality_profile(df_clean)
print("\n=== Perfil de calidad (DESPUÉS - df_clean) Top 15 por % nulos ===")
print(profile_after.head(15))

print("\n=== Muestra df_agg (lista para Bloques 3-5: integración/analítica/visualización) ===")
print(df_agg.head(10))

```

ANEXO C: Código en Python Fase Integración de datos

```

# =====
# BLOQUE 3: Integración de Datos (VERSIÓN ACTUALIZADA y COMPATIBLE)
# -----
# Elementos del bloque (modelo):
# (1) Mapeo semántico y normalización
# (2) Transformación y unificación
# (3) Modelado de datos integrados (vista analítica)
# (4) Resolución de conflictos
# (5) Gestión de metadatos
# (6) Validación semántica (reglas de negocio)
#
# Dataset: SIS Atenciones (2017-2020)
# Entrada: df_clean (salida del Bloque 2)
# Salidas:

```

```

# - df_integrated: vista integrada (hechos agregados) + Llave
# - data_dictionary: metadatos (diccionario básico)
# - qc_integracion: validaciones semánticas (conteos)
# - kpi_integracion: KPIs de integración (trazabilidad del bloque)
# =====

import pandas as pd
import numpy as np
import re
import hashlib
from typing import Optional

# -----
# 0) Pre-requisitos
# -----
assert "df_clean" in globals(), "No encuentro df_clean. Ejecuta primero el Bloque 2."
df_in = df_clean.copy()

pd.set_option("display.max_columns", 200)
pd.set_option("display.width", 200)

# -----
# Helpers (reutilizables)
# -----
def normalize_text(s: pd.Series) -> pd.Series:
    """
    Normaliza texto: strip, espacios múltiples, mayúsculas.
    Mantiene NaN como NaN (no lo convierte en 'NaN').
    """
    s2 = s.astype("string")
    s2 = s2.str.strip()
    s2 = s2.str.replace(r"\s+", " ", regex=True)
    s2 = s2.str.upper()
    return s2

```

```

def normalize_code(s: pd.Series, pad: Optional[int] = None) ->
pd.Series:
    """
    Normaliza códigos: strip; opcionalmente padding con ceros a la
    izquierda.
    Compatible con Python < 3.10 (usa Optional[int] en lugar de int |
    None).
    """
    s2 = s.astype("string").str.strip()
    if pad is not None:
        s2 = s2.str.zfill(pad)
    return s2

def build_key_hash(df: pd.DataFrame, cols: list, sep: str = "|") ->
pd.Series:
    """Construye llave hash estable (SHA1 corto) para
    trazabilidad/joins."""
    raw = df[cols].astype("string").fillna("").agg(sep.join, axis=1)
    return raw.apply(lambda x: hashlib.sha1(x.encode("utf-
    8")).hexdigest()[:12])

def make_data_dictionary(df: pd.DataFrame, sample_n: int = 3) ->
pd.DataFrame:
    """Diccionario básico: tipo, nulos, unicos, ejemplos."""
    rows = []
    for c in df.columns:
        s = df[c]
        ex = s.dropna().astype("string").head(sample_n).tolist()
        rows.append({
            "columna": c,
            "tipo": str(s.dtype),
            "nulos": int(s.isna().sum()),
            "unicos": int(s.nunique(dropna=True)),
            "ejemplos": "; ".join(ex) if ex else ""
        })
    return pd.DataFrame(rows)

```

```

# =====
# 1) Mapeo semántico y normalización
# =====

# 1.1 Renombrado estándar (defensivo)
rename_map = {
    "AÑO": "ANIO",
    "DESC_SERVICIO": "SERVICIO",
}
df = df_in.rename(columns={k: v for k, v in rename_map.items() if k in
df_in.columns}).copy()

# 1.2 Normalización de texto (reduce conflictos semánticos por
mayúsculas/espacios)
text_cols = [c for c in ["REGION", "PROVINCIA", "DISTRITO", "SERVICIO",
"SEXO", "GRUPO_EDAD", "IPRESS", "NIVEL_EESS"] if c in df.columns]
for c in text_cols:
    df[c] = normalize_text(df[c])

# 1.3 Normalización de códigos (integración semántica)
# UBIGEO generalmente es 6 dígitos; padding a 6 para estandarizar.
if "UBIGEO_DISTRITO" in df.columns:
    df["UBIGEO_DISTRITO"] = normalize_code(df["UBIGEO_DISTRITO"],
pad=6)

code_cols = [c for c in ["COD_SERVICIO", "COD_IPRESS"] if c in
df.columns]
for c in code_cols:
    df[c] = normalize_code(df[c])

# =====
# 2) Transformación y unificación (tipos + consistencias)
# =====

# 2.1 Tipos numéricos
for c in ["ANIO", "MES", "ANIO_ARCHIVO"]:
    if c in df.columns:

```

```

df[c] = pd.to_numeric(df[c], errors="coerce")

if "ATENCIONES" in df.columns:
    df["ATENCIONES"] = pd.to_numeric(df["ATENCIONES"], errors="coerce")

# 2.2 Resolución de conflictos
conflicts_anio = 0
if {"ANIO", "ANIO_ARCHIVO"}.issubset(df.columns):
    mask = df["ANIO"].notna() & df["ANIO_ARCHIVO"].notna() &
(df["ANIO"].astype("Int64") != df["ANIO_ARCHIVO"].astype("Int64"))
    conflicts_anio = int(mask.sum())
    # Política: priorizar ANIO_ARCHIVO (trazabilidad por origen)
    df.loc[mask, "ANIO"] = df.loc[mask, "ANIO_ARCHIVO"]

# MES válido (1..12)
invalid_mes = 0
if "MES" in df.columns:
    invalid_mes = int((~df["MES"].between(1, 12,
inclusive="both")).sum())
    df = df[df["MES"].between(1, 12, inclusive="both")].copy()

# ATENCIONES no negativas y no nulas
invalid_atenciones_neg = 0
invalid_atenciones_nan = 0
if "ATENCIONES" in df.columns:
    invalid_atenciones_nan = int(df["ATENCIONES"].isna().sum())
    df = df.dropna(subset=["ATENCIONES"]).copy()
    invalid_atenciones_neg = int((df["ATENCIONES"] < 0).sum())
    df = df[df["ATENCIONES"] >= 0].copy()
    df["ATENCIONES"] = df["ATENCIONES"].astype(np.int64)

# =====
# 3) Modelado de datos integrados (vista analítica)
# =====

grain = [c for c in ["ANIO", "MES", "REGION", "PROVINCIA", "SERVICIO",
"SEXO", "GRUPO_EDAD"] if c in df.columns]

```

```

# Opcional: incluir DISTRITO (más granular)
# if "DISTRITO" in df.columns:
#     grain.append("DISTRITO")

if len(grain) < 4:
    raise ValueError(f"Grano insuficiente para una vista integrada.
grain actual: {grain}")

df_integrated = (
    df.groupby(grain, dropna=False, as_index=False)
    .agg(ATENCIONES=("ATENCIONES", "sum"))
)

# Llaves de integración:
df_integrated["KEY_RAW"] =
df_integrated[grain].astype("string").fillna("").agg("|".join, axis=1)
df_integrated["KEY_INTEGRACION"] = build_key_hash(df_integrated, grain)

# =====
# 4) Gestión de metadatos (diccionario + dominios)
# =====

data_dictionary = make_data_dictionary(df_integrated, sample_n=3)

domain_summary = {}
for c in ["REGION", "PROVINCIA", "SERVICIO", "SEXO", "GRUPO_EDAD"]:
    if c in df_integrated.columns:
        domain_summary[c] = int(df_integrated[c].nunique(dropna=True))
domain_summary_df = pd.DataFrame([domain_summary]) if domain_summary
else pd.DataFrame()

# =====
# 5) Validación semántica (reglas de negocio)
# =====

rules = {}

```

```

rules["atenciones_no_negativas_violaciones"] =
int((df_integrated["ATENCIONES"] < 0).sum())

sexo_fuera = None
valores_sexo_fuera = None
if "SEXO" in df_integrated.columns:
    dominio_sexo = {"FEMENINO", "MASCULINO"}
    mask = ~df_integrated["SEXO"].isin(dominio_sexo)
    sexo_fuera = int(mask.sum())
    valores_sexo_fuera = df_integrated.loc[mask,
"SEXO"].value_counts().head(10).to_dict()

rules["sexo_fuera_de_dominio_count"] = sexo_fuera
rules["sexo_fuera_de_dominio_top"] = valores_sexo_fuera

anio_fuera = None
if "ANIO" in df_integrated.columns:
    anio_fuera = int((~df_integrated["ANIO"].between(2017, 2020,
inclusive="both")).sum())
rules["anio_fuera_rango_2017_2020_count"] = anio_fuera

mes_fuera = None
if "MES" in df_integrated.columns:
    mes_fuera = int((~df_integrated["MES"].between(1, 12,
inclusive="both")).sum())
rules["mes_fuera_1_12_count"] = mes_fuera

qc_integracion = pd.DataFrame([rules])

# =====
# 6) KPIs de integración (para el capítulo de resultados)
# =====

kpis = {
    "filas_df_clean": int(len(df_clean)),
    "columnas_df_clean": int(df_clean.shape[1]),
    "filas_post_unificacion": int(len(df)),

```

```

    "filas_integradas": int(len(df_integrated)),
    "columnas_integradas": int(df_integrated.shape[1]),
    "reduccion_filas_integracion_pct": round(100 * (1 -
(len(df_integrated) / max(len(df), 1))), 4),
    "grain_usado": ", ".join(grain),
    "conflictos_anio_detectados": int(conflicts_anio),
    "mes_invalidos_detectados": int(invalid_mes),
    "atenciones_nan_detectadas": int(invalid_atenciones_nan),
    "atenciones_neg_detectadas": int(invalid_atenciones_neg),
}

kpi_integracion = pd.DataFrame([kpis])

print("=== BLOQUE 3 (Integración) - KPIs ===")
display(kpi_integracion)

print("\n=== BLOQUE 3 (Integración) - Validación semántica ===")
display(qc_integracion)

if not domain_summary_df.empty:
    print("\n=== BLOQUE 3 (Integración) - Resumen de dominios
(catálogos) ===")
    display(domain_summary_df)

print("\n=== BLOQUE 3 (Integración) - Muestra df_integrated ===")
display(df_integrated.head(10))

print("\n=== Diccionario de datos (integrado) ===")
display(data_dictionary.head(30))

print("\n✅ Bloque 3 completado. Dataset integrado listo para Bloque
4.")
print("Filas/Columnas (df_integrated):", df_integrated.shape)

```

ANEXO D: Código en Python Fase Analítica Avanzada

```
# =====
# BLOQUE 4: Capacidad Analítica Avanzada (VERSION ESTRUCTURADA)
# -----
# 4.1 Analítica Descriptiva -> ¿Qué pasó?
# 4.2 Analítica Predictiva -> ¿Qué pasará?
# 4.3 Analítica Prescriptiva -> ¿Qué hacer?
# Entrada: df_integrated (Bloque 3)
# =====

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.metrics import mean_absolute_error, mean_squared_error
from sklearn.ensemble import RandomForestRegressor

pd.set_option("display.max_columns", 200)
pd.set_option("display.width", 200)

# -----
# 0) Pre-requisitos
# -----
assert "df_integrated" in globals(), "No encuentro df_integrated.
Ejecuta primero el Bloque 3."
dfA = df_integrated.copy()

for c in ["ANIO", "MES", "ATENCIONES"]:
    if c not in dfA.columns:
        raise ValueError(f"Falta columna requerida: {c}")

dfA["ANIO"] = pd.to_numeric(dfA["ANIO"],
errors="coerce").astype("Int64")
```

```

dfa["MES"] = pd.to_numeric(dfa["MES"], errors="coerce").astype("Int64")
dfa["ATENCIONES"] = pd.to_numeric(dfa["ATENCIONES"],
errors="coerce").fillna(0).astype(np.int64)

dfa = dfa.dropna(subset=["ANIO", "MES"]).copy()
dfa["DATE"] = pd.to_datetime(dfa["ANIO"].astype(str) + "-" +
dfa["MES"].astype(str) + "-01", errors="coerce")
dfa = dfa.dropna(subset=["DATE"]).copy()
dfa["MONTH_INDEX"] = (dfa["ANIO"] - dfa["ANIO"].min()) * 12 +
(dfa["MES"] - 1)

# =====
# 4.1 ANALÍTICA DESCRIPTIVA (¿Qué pasó?)
# =====

# T4_1A: KPIs globales
T4_1A_kpis = pd.DataFrame([{
    "periodo_inicio": dfa["DATE"].min().date(),
    "periodo_fin": dfa["DATE"].max().date(),
    "total_atenciones": int(dfa["ATENCIONES"].sum()),
    "promedio_atenciones_por_registro":
float(dfa["ATENCIONES"].mean()),
    "n_registros_analiticos": int(len(dfa)),
    "n_regiones": int(dfa["REGION"].nunique()) if "REGION" in
dfa.columns else None,
    "n_servicios": int(dfa["SERVICIO"].nunique()) if "SERVICIO" in
dfa.columns else None
}])

print("=== 4.1 DESCRIPTIVA: Cuadro T4_1A KPIs globales ===")
display(T4_1A_kpis)

# T4_1B: Tendencia mensual (agregada)
T4_1B_trend = (
    dfa.groupby("DATE", as_index=False)
        .agg(ATENCIONES=("ATENCIONES", "sum"))
        .sort_values("DATE")

```

```

)
T4_1B_trend["variacion_pct_mensual"] =
T4_1B_trend["ATENCIONES"].pct_change() * 100

print("\n=== 4.1 DESCRIPTIVA: Cuadro T4_1B Tendencia mensual ===")
display(T4_1B_trend)

# Figura: Tendencia mensual
plt.figure()
plt.plot(T4_1B_trend["DATE"], T4_1B_trend["ATENCIONES"])
plt.title("Tendencia mensual de atenciones (total)")
plt.xlabel("Fecha")
plt.ylabel("Atenciones")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# T4_1C: Segmentación por sexo
T4_1C_sexo = None
if "SEXO" in dfA.columns:
    T4_1C_sexo = (
        dfA.groupby("SEXO", as_index=False)
            .agg(ATENCIONES=("ATENCIONES", "sum"))
            .sort_values("ATENCIONES", ascending=False)
    )
    print("\n=== 4.1 DESCRIPTIVA: Cuadro T4_1C Segmentación por sexo
===")
    display(T4_1C_sexo)

# T4_1D: Segmentación por grupo de edad
T4_1D_edad = None
if "GRUPO_EDAD" in dfA.columns:
    T4_1D_edad = (
        dfA.groupby("GRUPO_EDAD", as_index=False)
            .agg(ATENCIONES=("ATENCIONES", "sum"))
            .sort_values("ATENCIONES", ascending=False)
    )

```

```

    print("\n=== 4.1 DESCRIPTIVA: Cuadro T4_1D Segmentación por grupo
de edad ===")
    display(T4_1D_edad)

# =====
# 4.2 ANALÍTICA PREDICTIVA (¿Qué pasará?)
# =====

target = "ATENCIONES"
cat_cols = [c for c in ["REGION", "PROVINCIA", "SERVICIO", "SEXO",
"GRUPO_EDAD"] if c in dfA.columns]
num_cols = ["MONTH_INDEX"]

X = dfA[cat_cols + num_cols].copy()
y = dfA[target].copy()

# Split temporal si existe 2020 (entrena 2017-2019, prueba 2020)
use_temporal_split = (dfA["ANIO"].min() <= 2019 and dfA["ANIO"].max()
>= 2020)

if use_temporal_split:
    train_mask = dfA["ANIO"] <= 2019
    X_train, y_train = X[train_mask], y[train_mask]
    X_test, y_test = X[~train_mask], y[~train_mask]
    split_name = "temporal_2017_2019_vs_2020"
else:
    X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
    split_name = "random_80_20"

preprocess = ColumnTransformer(
    transformers=[
        ("cat", OneHotEncoder(handle_unknown="ignore"), cat_cols),
        ("num", "passthrough", num_cols),
    ],
    remainder="drop"
)

```

```

pred_model = RandomForestRegressor(
    n_estimators=200,
    random_state=42,
    n_jobs=-1
)

pipeline = Pipeline(steps=[("prep", preprocess), ("model",
pred_model)])
pipeline.fit(X_train, y_train)

y_pred = pipeline.predict(X_test)

mae = mean_absolute_error(y_test, y_pred)
rmse = mean_squared_error(y_test, y_pred, squared=False)

# T4_2A: Métricas del modelo
T4_2A_pred_metrics = pd.DataFrame([
    "split": split_name,
    "MAE": float(mae),
    "RMSE": float(rmse),
    "n_train": int(len(X_train)),
    "n_test": int(len(X_test)),
    "features_categoricas": ", ".join(cat_cols),
    "feature_tiempo": "MONTH_INDEX"
])

print("\n=== 4.2 PREDICTIVA: Cuadro T4_2A Métricas del modelo ===")
display(T4_2A_pred_metrics)

# T4_2B: Real vs Predicho (muestra)
T4_2B_real_vs_pred = X_test.copy()
T4_2B_real_vs_pred["ATENCIONES_REAL"] = y_test.values
T4_2B_real_vs_pred["ATENCIONES_PRED"] = np.round(y_pred).astype(int)
T4_2B_real_vs_pred["ERROR_ABS"] =
(T4_2B_real_vs_pred["ATENCIONES_REAL"] -
T4_2B_real_vs_pred["ATENCIONES_PRED"]).abs()

```

```

T4_2B_real_vs_pred = T4_2B_real_vs_pred.sort_values("ERROR_ABS",
ascending=False).head(20)

print("\n=== 4.2 PREDICTIVA: Cuadro T4_2B Real vs Predicho (Top 20 por
error) ===")
display(T4_2B_real_vs_pred)

# =====
# 4.3 ANALÍTICA PRESCRIPTIVA (¿Qué hacer?)
# =====

last_date = dfA["DATE"].max()
next_month_index = int(dfA["MONTH_INDEX"].max() + 1)

# Tomar combinaciones más frecuentes para escenario (control de
dimensionalidad)
base_cols = cat_cols.copy()
top_combos = (
    dfA.groupby(base_cols, as_index=False)
        .agg(freq=("ATENCIONES", "count"),
atenciones_total=("ATENCIONES", "sum"))
        .sort_values(["freq", "atenciones_total"], ascending=False)
        .head(500)
        .drop(columns=["freq", "atenciones_total"])
)

scenario_next = top_combos.copy()
scenario_next["MONTH_INDEX"] = next_month_index
scenario_next["ATENCIONES_PRED_NEXT"] =
pipeline.predict(scenario_next[base_cols +
["MONTH_INDEX"]]).clip(min=0)

p75 = float(np.percentile(scenario_next["ATENCIONES_PRED_NEXT"], 75))
p90 = float(np.percentile(scenario_next["ATENCIONES_PRED_NEXT"], 90))

def priority_label(x):

```

```

    if x >= p90:
        return "ALTA"
    if x >= p75:
        return "MEDIA"
    return "BAJA"

scenario_next["PRIORIDAD"] =
scenario_next["ATENCIONES_PRED_NEXT"].apply(priority_label)

# T4_3A: Recomendaciones (Top 30 prioridad ALTA)
T4_3A_recommendations = (
    scenario_next[scenario_next["PRIORIDAD"] == "ALTA"]
    .sort_values("ATENCIONES_PRED_NEXT", ascending=False)
    .head(30)
)

# texto breve de recomendación
if "SERVICIO" in T4_3A_recommendations.columns and "REGION" in
T4_3A_recommendations.columns and "PROVINCIA" in
T4_3A_recommendations.columns:
    T4_3A_recommendations["RECOMENDACION"] = (
        "Priorizar recursos para " +
T4_3A_recommendations["SERVICIO"].astype(str)
        + " en " + T4_3A_recommendations["REGION"].astype(str)
        + " (Prov. " + T4_3A_recommendations["PROVINCIA"].astype(str) +
    ")."
    )

print("\n=== 4.3 PRESCRIPTIVA: Cuadro T4_3A Recomendaciones (Top 30
prioridad ALTA) ===")
display(T4_3A_recommendations)

print("\n☑ Bloque 4 completado: Descriptiva (T4_1*), Predictiva
(T4_2*), Prescriptiva (T4_3*).")

```