# Optimización de la Gestión de Datos en Salud de la Clínica Los Cobos: Fomentando la Seguridad y el Uso Responsable de la Información

Paula Andrea Castellanos Tabio, Angie Stefani Caro Ballestas, Linda Tábatha Martínez Lara, Diana Patricia Caro Ballestas y Tania Lucia Diaz Alfaro

Especialización en Big Data, Fundación Universitaria Compensar
Propuesta de Consultoría

Docente: Wilson Hernando Soto Urrea

27 de mayo de 2024

Esta obra está bajo una Licencia Creative Commons Atribución-NoComercialCompartirIgual 4.0 Internacional.

# Optimización de la Gestión de Datos en Salud de la Clínica Los Cobos: Fomentando la Seguridad y el Uso Responsable de la Información

**Paula Andrea Castellanos Tabio, Angie Stefani Caro Ballestas, Linda Tábatha Martínez Lara, Diana Patricia Caro Ballestas y Tania Lucia Diaz Alfaro**

Trabajo de grado presentado como requisito parcial para optar al título de:
**Especialista en Big Data**

Director(a):
Docente: Wilson Hernando Soto Urrea

Fundación Universitaria Compensar
Facultad de Ingeniería, Especialización en Big Data
Bogotá, Colombia
2024

# TABLA DE CONTENIDO

1. Introducción....4
1.1. Antecedentes y Justificación....4
1.2. Formulación del problema....5
2. Objetivos de la investigación....6
3. Marco teórico....6
4. Diseño Metodológico....8
5. La Gobernanza de Datos como aspecto fundamental en la generación de valor de la Organización....10
Conclusiones....13
Referencias....15

# 1. Introducción

Actualmente, el uso y el análisis a partir del Big Data se ha globalizado ya que las organizaciones han almacenado datos a lo largo de su trayectoria, las herramientas ofimáticas han sido muy útiles para almacenar información, y a medida que se ha avanzado tecnológicamente, se han desarrollado nuevos programas para facilitar no solo el almacenamiento, sino también el análisis y la visualización de los datos masivos.

Los avances del segundo decenio del siglo XXI facilitan el análisis de Big Data, pero las compañías aún enfrentan grandes desafíos en la recolección de datos, entre estos la falta de procedimientos y estrategias claras para la gobernabilidad de los datos hacen que la estructura de estos no sea eficiente y eficaz para tomar decisiones.

La gobernanza de datos se refiere a la organización de sistemas, políticas, estándares y controles que garantizan la calidad, integridad, seguridad y disponibilidad de los datos en toda la organización. (Manar, s.f.). En otras palabras, se trata de crear un sistema que maximice el valor de los datos, minimice el riesgo y garantice el cumplimiento de las regulaciones y estándares existentes. En este caso, la gestión de datos abarca aspectos importantes como la definición de servicios y funciones, la gestión de la privacidad y seguridad, la calidad de los datos, la detección y comprensión en el uso de la información y, además implica la implementación de tecnología.

En este sentido, una sólida gestión de datos ayuda a mejorar la toma de decisiones y el desempeño de diversas organizaciones, y genera confianza en el equipo de trabajo, en los clientes y garantiza el cumplimiento de los requisitos legales y normatividad.

## 1.1. Antecedentes y Justificación

Según la publicación Hackeo masivo en Colombia: "La información de millones de personas está en manos de delincuentes en este momento", este comunicado fue expuesto por el Periódico El País el 14 de septiembre del 2023, resaltando un incidente en el cual los datos de diversas entidades gubernamentales, entre estas el Ministerio de Salud, junto con otras del sector privado, fueron víctimas de un ciberataque, consistente en el secuestro de los datos alojados en máquinas virtuales de IFX Networks, a través de un ramsomware, arriesgaron y vulneraron la información de sus clientes. Esta situación lleva a considerar la importancia de las normas vigentes para proteger datos sensibles y los procesos definidos al interior de las organizaciones para prevenir posibles ciberataques que pongan en riesgo la seguridad de la información.

Otro caso reciente de igual relevancia es el que presentó la Contraloría General de la Nación en el comunicado de prensa No.125, del 3 de agosto de 2023, en el que se menciona la auditoría a la empresa prestadora de salud Emssanar, que presentó hallazgos, algunos de carácter fiscal, en donde sobresale un problema de pagos indebidos por la falta de actualización de los datos de los afiliados.

Casos como los presentados, dejan vislumbrar el problema de algunas empresas del sector salud en Colombia al gestionar datos transaccionales o relevantes para tomar decisiones, errores tales como duplicidad de la información, falta de consistencia o información incomprensible, deriva en ausencia de credibilidad, exactitud y no poder realizar trazabilidad de la información.

Las organizaciones especialmente del sector salud tienen un reto importante asociado a la comprensión y definición de los diferentes datos que se generan constantemente, de acuerdo con este panorama se considera establecer una solución con la aplicación de la guía técnica de gobernanza de datos para el catálogo de productos farmacéuticos con la intención de conocer y definir el ciclo de vida de los datos, identificar los custodios o personas encargadas de garantizar el adecuado manejo de estos y finalmente quienes serán los consumidores finales de estos datos.

La gobernanza del dato en la creación de productos farmacéuticos en el sistema de información inicia con la identificación de los datos propios a intervenir, analizar el estado actual y realizar el correcto perfilamiento, continuando con la definición e implementación de la guía técnica.

### 1.2.Formulación del problema

El enfoque de la problemática se concentra en la empresa “Los Cobos Medical Center” que pertenece a la industria de la salud, se caracteriza por la prestación de servicios de salud integral, con un modelo de atención innovador, centrado en el paciente y su familia, donde la humanización del servicio, el conocimiento médico científico y académico se integran en una cohesión de protección y seguimiento de la salud de sus pacientes.

Actualmente la organización en referencia enfrenta desafíos en la gestión y control de los datos, específicamente en el proceso asociado a la creación del dato maestro en el sistema de información - SAP de productos farmacéuticos porque la ausencia de buenas prácticas en el uso de los datos ha desencadenado duplicidad y falta de información, acceso no seguro a los datos, deficiencias en el costo médico que sin duda impacta en la atención segura al paciente, uso

eficiente de recursos, control de inventarios especialmente relacionado con productos farmacéuticos y suministro oportuno y eficiente de medicamentos. De acuerdo con lo anterior, surge la necesidad de generar soluciones orientadas a la implementación de buenas prácticas y correcta gobernabilidad de los datos, obteniendo como resultado la recolección, gestión y control de los datos de manera adecuada permitiendo a la organización tener una estructura acorde con los pilares de la seguridad de la información que se caracteriza por la confidencialidad, integridad y disponibilidad de los datos.

## 2. Objetivos de la investigación

### Objetivo General

Establecer un proceso de gestión de los datos en una organización de salud, que fomente el uso responsable y la seguridad de la información.

### Objetivos Específicos

- Implementar un procedimiento que mejore la validez, confiabilidad y pertinencia de los datos para su uso específico.
- Aplicar una herramienta que posibilite el mejoramiento significativamente de la gestión de datos sensibles y confidenciales, restringiendo accesos no autorizados y previniendo el daño o pérdida de información valiosa.
- Buscar la eficacia en la disponibilidad y accesibilidad de los datos según sea necesario, priorizando la protección y la integridad de estos.

## 3. Marco teórico

### 3.1. Gobernanza del dato

De acuerdo con el interés de este proyecto encaminado a la presentación de una propuesta de gobernanza del dato específicamente para productos farmacéuticos, se ha considerado relevante que esta comprende “Un conjunto de prácticas para tomar decisiones acerca de los datos y gestionarlos durante todo su ciclo de vida, con el objeto de optimizar la capacidad de la organización para utilizar datos que generen información de calidad a fin de fundamentar políticas, estrategias y gestión operativa.” (Organización Panamericana de la salud - Organización Mundial de la Salud, 2019), lo anterior sujeto a la capacidad que tengan las instituciones de salud para disponer de tiempo y recursos que impacten en el fortalecimiento y mejora continua de la gobernanza de datos.

La gobernanza del dato se acompaña de elementos fundamentales relacionados entre sí, como Big Data y metodologías para la implementación de proyectos asociados a Big Data – KDD, SEMMA y CRISP–DM.

### 3.2. Big Data

El término “Big Data” abarca datos que contienen una mayor variedad y que se presentan en volúmenes crecientes y a una velocidad superior. El big data está formado por conjuntos de datos de mayor tamaño y más complejos, especialmente procedentes de nuevas fuentes de datos. (ORACLE, s.f.)

### 3.3. KDD (Knowledge Discovery in Database)

Es un proceso interactivo y recursivo para descubrir patrones y relaciones valiosas en grandes conjuntos de datos, este proceso involucra varios pasos: Comprender el dominio de la aplicación / Extraer la base de datos objetivo / preparar los datos / Minería de Datos / Interpretación / Utilizar el conocimiento descubierto. (Riquelme, Ruiz, & Gilbert, 2006)

### 3.4. SEMMA (Sample, Explore, Modify, Model, Assess)

Esta metodología fue propuesta por SAS Institute, el cual la define como el proceso de selección, exploración y modelado aplicado a cantidades significativas de datos almacenados que permitan el descubrimiento de patrones como herramientas de apoyo para el negocio. Según SAS más que una metodología para minería de datos, SEMMA es un conjunto de herramientas funcionales enfocadas hacia los aspectos de desarrollo propio de un modelo de minería. (Hernández & Dueñas, 2009). Las fases que componen la metodología en referencia son: muestreo, exploración, modificación, modelado y evaluación.

### 3.5. CRISP-DM (CRoss-Industry Standard Process for Data Mining)

Esta metodología fue diseñada en 1999 por un consorcio de empresas europeas (NCR (Dinamarca), AG (Alemania), SPSS (Inglaterra) y OHRA (Holanda)), los cuales unieron sus esfuerzos para crear un esquema que permitiera mostrar el ciclo de vida de un proyecto de minería de datos. (Hernández & Dueñas, 2009). Las fases que componen la metodología en referencia son: Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación, Despliegue, Monitorización.

Desde la comprensión del negocio hasta la implementación y el monitoreo de modelos en producción, estas metodologías brindan un marco estructurado para abordar proyectos de minería

de datos. El contexto y las necesidades particulares del proyecto determinan la metodología a utilizar.

## 4. Diseño Metodológico

A continuación, se comparan tres metodologías útiles para solucionar el problema de validación de la base de datos de productos farmacéuticos.

Según (Moine, Haedo, & Gordillo, 2011) en su estudio comparativo de metodologías para minería de datos, la minería de datos es el proceso de extraer conocimiento útil, comprensible y novedoso de grandes volúmenes de datos. Teniendo el documento mencionado se elaboró en cuadro comparativo que registra las diferencias de las metodologías CRISP-DM, SEMMA y KDD.

***Tabla 1 – Comparativo Metodologías para Minería de Datos***

| Metodología | CRISP-DM | SEMMA | KDD |
|---|---|---|---|
| **Descripción** | Proceso estándar para proyectos de minería de datos. Incluye seis fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue. | Proceso desarrollado por SAS Institute. Incluye cinco fases: muestra, exploración, modificación, modelado y evaluación. | Proceso de descubrimiento de conocimiento en bases de datos. Incluye selección, preprocesamiento, transformación, minería de datos y evaluación. |
| **Aplicación** | Comprender el contexto del problema, identificar datos disponibles, desarrollar modelos predictivos o de detección de anomalías. | Seleccionar muestras representativas, explorar patrones en los datos, construir modelos predictivos o descriptivos. | Seleccionar datos relevantes, preprocesar y limpiar datos, aplicar algoritmos de minería. |
| **Ventajas** | Estructura clara y paso a paso. Ampliamente utilizado. | Iterativo y adaptable. Enfoque práctico. Fácil de implementar en proyectos cortos. | Marco general para el descubrimiento de conocimiento. Flexible. Profunda comprensión de los datos. |
| **Desventajas** | Requiere tiempo y recursos significativos. Puede ser rígido en proyectos ágiles. Compleja para equipos pequeños. | Menos estructurado que CRISP-DM. Requiere experiencia en herramientas específicas de SAS. | Complejo de implementar. Menos estructurado que CRISP-DM en fases específicas. A veces demasiado teórico. |

***Fuente – Creación propia***

De acuerdo con la descripción de la ***Tabla 1*** aunque KDD y SEMMA pueden ser útiles en ciertos contextos, no son las metodologías más adecuadas para solucionar el problema de validación de la base de datos de productos farmacéuticos debido a lo siguiente:

**Enfoque limitado:** Tanto KDD como SEMMA se enfocan principalmente en el proceso de minería de datos, dejando de lado aspectos importantes como la comprensión del negocio y el monitoreo continuo.

**Falta de iteratividad:** Aunque KDD y SEMMA tienen etapas definidas, carecen de la iteratividad y flexibilidad necesarias para abordar problemas de roles, perfiles de acceso y control en la creación de productos farmacéuticos.

**Énfasis en la minería de datos:** Ambas metodologías se centran en la aplicación de técnicas de minería de datos, pero no ponen suficiente énfasis en la calidad de los datos y en la comprensión del negocio. Esto puede llevar a soluciones que no abordan adecuadamente los problemas de duplicidad, ausencia de información y datos desactualizados.

**Falta de enfoque en la aplicación práctica:** KDD y SEMMA no tienen un enfoque claro en la aplicación práctica de los resultados obtenidos a través de la minería de datos. Esto puede dificultar la implementación de soluciones concretas para mitigar los riesgos directos en la atención y seguridad de los pacientes.

Por lo anterior, la metodología que más se ajusta a la problemática descrita es CRISP-DM (Cross-Industry Standard Process for Data Mining) - "Proceso Estándar Inter-Industrias para Minería de Datos", por lo siguiente:

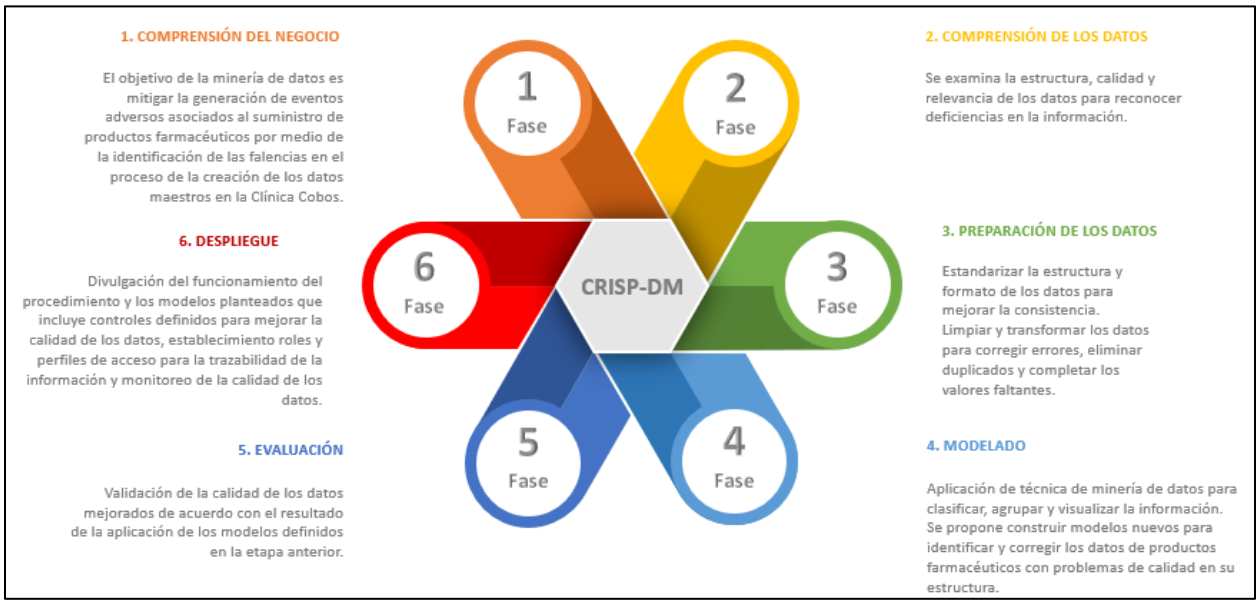
**Estructura y Flexibilidad:** CRISP-DM es una metodología estructurada y flexible que se adapta a diferentes tipos de problemas, incluyendo la validación y limpieza de datos. Permite un enfoque paso a paso para abordar los desafíos identificados en la base de datos de productos farmacéuticos.

**Fases Iterativas:** CRISP-DM se compone de seis fases iterativas:

- Comprensión del negocio.
- Comprensión de los datos.
- Preparación de los datos.
- Modelado.
- Evaluación.
- Despliegue.

Estas fases permiten abordar de manera sistemática los problemas de duplicidad, ausencia de información, datos desactualizados, falta de estandarización y control en la creación de productos farmacéuticos.

*Figura 1 - Propuesta de solución basada en la metodología - CRISP-DM*




*Fuente – Creación propia*

## 5. La Gobernanza de Datos como aspecto fundamental en la generación de valor de la Organización

El impacto de la gobernanza de datos en las instituciones de salud recae en la calidad de la atención médica, la seguridad del paciente, el uso adecuado de recursos y la toma de decisiones pertinentes soportadas en información. Además, la transferencia del conocimiento como resultado de las buenas prácticas asociadas al uso de los datos promueve la construcción de estándares y políticas para su gestión garantizada respecto a su calidad y coherencia.

Otro aspecto relevante en la gobernanza de los datos es la integración de los datos que genera conocimiento a partir de información completa y consistente para tomar decisiones clínicas y administrativas, además contribuye al desarrollo de proyectos de investigación que se pueden adelantar al interior de la institución de salud y garantiza la disponibilidad de datos cubiertos por un entorno que protege la privacidad y seguridad de la información.

Al realizar el análisis de la base de datos maestros de la organización objeto de estudio, se observa que los campos no están diligenciados en su totalidad lo que genera falta de información en el proceso, a continuación, se detallan las columnas y la cantidad de espacios en blanco:

***Tabla 2 – Exploración de la base de datos***

| Item | Conteo de Blancos |
|---|---|
| Alto riesgo prestación | 7.516 |
| Radiofármaco | 7.516 |
| Confirmación de uso | 7.515 |
| Antibiotico Uso Restringido | 7.514 |
| Electrolito concentrado | 7.509 |
| Alto riesgo LASA | 7.475 |
| Medicamento control especial Prestacion | 7.448 |
| Medicamento control especial Medication | 7.441 |
| Oncológico | 7.410 |
| Medicamento requiere adecuación | 7.393 |
| Antibiotico | 7.389 |
| Multidosis | 7.318 |
| Alto Riesgo Farmacológico Medication | 7.286 |
| Medicamento estable | 7.241 |
| Status Material Centro | 7.198 |
| B.prestación | 7.187 |
| IUM Material | 6.840 |
| No PBS medication | 6.795 |
| Código grupo terapéutico | 6.538 |
| Descripción de grupo terapéutico | 6.538 |
| Codigo Subgrupo Terapéutico | 6.537 |
| Descripción Subgrupo Terapéutico | 6.537 |
| ATC Medication | 6.535 |
| Principio Activo | 6.533 |
| Ambito | 6.504 |
| Codigo Vias Administracion | 6.498 |
| Via de Administracion Medication | 6.498 |
| Codigo Forma Farmaceutica | 6.498 |
| Forma Farmaceutica | 6.498 |

| Item | Conteo de Blancos |
|---|---|
| Dosis Alternativa | 6.497 |
| Tamano envase | 6.497 |
| Descripción del medicamento | 6.496 |
| Dosis de Referencia | 6.496 |
| Unidad Dosis de Referencia | 6.496 |
| Unidad de Dosis Alternativa | 6.496 |
| Unidad de tamano de envase | 6.496 |
| Envase | 6.496 |
| Categoría del medicamento | 6.496 |
| ATC Material | 5.510 |
| Codigo Unico Medicamento Material | 4.407 |
| F.Farma/Riesgo | 2.181 |
| Prioridad de Adquisición | 738 |
| Código de Prestación | 407 |
| Descripcion de Prestación | 407 |
| Unidad de medida base de la prestación | 407 |
| No PBS Prestación | 407 |
| Régimen de control de precios | 407 |
| Acto administrativo cntrl prec | 407 |
| Fecha inicio control de precios | 407 |
| No Divisible | 407 |
| Código Genérico | 312 |
| Fecha de vencimiento del RS | 280 |
| Material sujeto a lote | 144 |
| Registro sanitario material | 121 |
| Clase de material | 64 |
| Segmentación | 56 |
| Descripcion de material | 5 |

***Fuente – Creación propia***

Con una totalidad de 510.808 datos, los campos en blanco son 279.245, representando el 54.7% de la totalidad. Una posible causa de este evento es que haya información que no aplique en todos los artículos, una buena práctica es diligenciar los campos con el símbolo “N/A”, indicando que ésta no es requerida para el mismo, asegurando que todo esté diligenciado y que no haya riesgo de pérdida de información, es importante resaltar información en blanco relevante como fechas de vencimiento, riesgo biológico, medicamento de control especial, entre otros.

Otro aspecto relevante detectado en la minería de los datos es la estructura. Dentro de la base de datos se evidencia que parte de la clasificación de los artículos se deja como una descripción de material, con caracteres alfanuméricos de escritura a percepción personal del digitador.

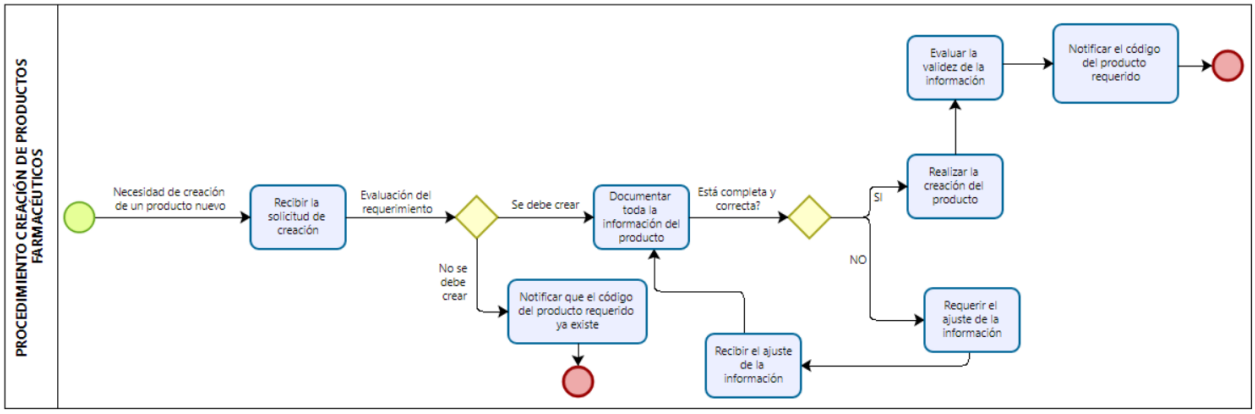
Un caso puntual se muestra a continuación, donde la diferencia de los artículos mostrados es el proveedor ya que en lugar de incluir un campo en la tabla se registra como texto en la descripción dificultando la clasificación correcta de los datos.

*Tabla 3 – Ejemplo estructura de clasificación de datos*

| Código Material | Texto breve material | Descripción de material |
|---|---|---|
| 21601000911 | ACETAMINOFEN-150MG/5ML/60ML-JBE-FCO-GEN | ACETAMINOFEN-150MG/5ML/60ML-JARABE-VIA ORAL-FCO-GENFAR |
| 21601001059 | ACETAMINOFEN-150MG-5ML-60ML-JBE-FCO-LAPR | ACETAMINOFEN-150MG-5ML-60ML-JARABE-VIA ORAL-FCO-LAPROFF SA |

*Fuente – Creación propia*

*Figura 2 – Procedimiento creación de productos farmacéuticos*




*Fuente – Creación propia*

## Conclusiones

El análisis de datos transciende en la generación de conocimiento, dependiendo de cómo se analizan los datos donde se incluya inteligencia artificial, aprendizaje automático y análisis avanzados que conllevan a identificar patrones y tendencias. Estos conocimientos pueden ser valiosos mejorando aspectos clínicos, administrativos y financieros de la institución.

Los principios clave para la creación de un proyecto de investigación son los siguientes: en primer lugar, la diversidad de fuentes de ideas sirve como punto de partida para el estudio de la organización; en segundo lugar, se requiere una definición precisa de los objetivos, asegurando que sean medibles y alcanzables; en tercer lugar, es crucial elegir el enfoque adecuado, ya sea cuantitativo o cualitativo; en cuarto lugar, se debe diseñar el estudio de manera apropiada; y finalmente, se debe evaluar la viabilidad del estudio en términos de recursos disponibles y plazos de ejecución.

Dentro de las metodologías de Big Data existen la KDD y SEMMA estas pueden ser útiles en ciertos contextos, sin embargo, no son las metodologías más adecuadas para solucionar el problema de validación de la base de datos de productos farmacéuticos debido a su enfoque limitado, falta de iteratividad, énfasis en la minería de datos y falta de enfoque en la aplicación práctica. En este caso, la metodología CRISP-DM es más apropiada por su estructura más amplia, enfoque en la calidad de los datos, iteratividad y énfasis en la aplicación práctica de los resultados.

La metodología CRISP-DM se enfoca en formular una estrategia adaptable a las condiciones de la minería de datos y la gestión del conocimiento, que se consideran principios complementarios para el crecimiento organizacional. El propósito es retroalimentar al negocio desde el conocimiento y la comprensión en el contexto de buenas prácticas para la gestión de los datos.

La integración de metodologías Big Data en las operaciones de una organización combinada con la colaboración de equipos interdisciplinarios, constituye un enfoque poderoso para abordar las necesidades y desafíos actuales de la organización, permitiendo identificar problemas de manera más eficiente y efectiva, lo que conlleva a tomar decisiones informadas en tiempo real. Este trabajo mancomunado facilita la aplicación de mejores prácticas y la adopción de herramientas tecnológicas innovadoras, además, fomenta la colaboración y el intercambio de

ideas entre diferentes áreas de la organización, promoviendo la innovación y el crecimiento a largo plazo.

La gobernanza de los datos mejora la disponibilidad de información segura, confiable y precisa al definir roles y perfiles para cada área involucrada en la generación de datos. Al establecer procedimientos internos que incluyan puntos de control, responsabilidades claras y tiempos de respuesta definidos, se pueden optimizar los procesos y contribuir a la reducción de tiempos y costos, además de identificar usuarios claves e información de calidad de acuerdo con la experticia técnica de los involucrados.

**Referencias**

Manar. (s.f.). *Gobernanza de Datos, ¿qué es?* Recuperado el 08 de abril de 2024 de https://manar.com.co/gobernanza-de-datos-que-es/

Organización Panamericana de la salud - Organización Mundial de la Salud. (2019). *IS4H TOOLKIT Cápsulas de conocimiento - Gobernanza de datos en salud pública*.

ORACLE. (s.f.). *¿Qué es big data?* Recuperado el 13 de mayo de 2024 de https://www.oracle.com/co/big-data/what-is-big-data/

Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de Datos: Conceptos y Tendencias. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, 11-18. https://www.redalyc.org/pdf/925/92502902.pdf

Hernández, C., & Dueñas, M. X. (2009). *Hacia una metodología de gestión del conocimiento basada en minería de datos [Archivo PDF].* http://repositorio.uigv.edu.pe/bitstream/handle/20.500.11818/982/COMTEL-2009-80-96.pdf?sequence=1&isAllowed=y

Moine, J. M., Haedo, A. S., & Gordillo, S. (2011). *Estudio Comparativo de Metodologías para Minería de Datos. [Archivo PDF]* https://sedici.unlp.edu.ar/bitstream/handle/10915/20034/Documento_completo.pdf?sequence=1&isAllowed=y