

Diseño de una solución de estandarización de fuentes de datos para optimizar la gestión de Información legal en ADRES

Jeison Oswaldo Bernal Vargas, Lina Mayerly Roa Rosas, Sandra Paola Benítez León, Mónica Yaneth Otavo Zamudio, Nataly Pérez Laiseca

Especialización de Big Data, Fundación Universitaria Compensar

Proyecto fin de grado

Omar Alexander León García

27 de noviembre de 2023



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

Diseño de una solución de estandarización de fuentes de datos para optimizar la gestión de Información legal en ADRES

Jeison Oswaldo Bernal Vargas, Lina Mayerly Roa Rosas, Sandra Paola Benítez León, Mónica Yaneth Otavo Zamudio, Nataly Pérez Laiseca

Trabajo de grado presentado como requisito parcial para optar al título de:

Especialización de Big Data

Director (a):

Título (PhD M.Sc., M.Eng, Ingeniero(a), etc.) y nombre del director(a)

Codirector (a):

Línea de Investigación:

Nombrar la línea de investigación en la que se enmarca el trabajo de investigación

Grupo de Investigación:

Nombrar el grupo en caso de que sea posible

Fundación Universitaria Unipanamericana – Compensar

Facultad de Ingeniería, Especialización en Big Data

Bogotá D.C., Colombia

2023

Resumen

La Administradora de los Recursos del Sistema General de Seguridad Social en Salud (ADRES) en Colombia enfrenta desafíos debido a la falta de estandarización de bases de datos, especialmente en el área de asesoría jurídica, generando dificultades de acceso a la información, inconsistencias y errores. Este proyecto propone utilizar la metodología CRISP-DM y la herramienta KNIME para estandarizar las fuentes de información en el proceso de reparto. Los objetivos incluyen determinar requerimientos, identificar fuentes de información y diseñar un proceso de estandarización. La metodología CRISP-DM guiará la comprensión del proceso y la preparación de datos, seguido por el modelado, evaluación y despliegue de soluciones garantizando la coherencia de los datos legales. Se evidencia que por medio del proceso de estandarización de datos se eliminan los registros duplicados, dando un formato igual a los nombres de los apoderados y a las fechas.

Palabras clave: ADRES, estandarización de datos, CRISP-DM, KNIME, eficiencia, toma de decisiones.

Abstract

The Administrator of the General System of Social Security Resources in Health (ADRES) in Colombia faces challenges due to the lack of database standardization, especially in the legal advisory area, leading to difficulties in accessing information, inconsistencies, and errors. This project proposes using the CRISP-DM methodology and the KNIME tool to standardize information sources in the distribution process. Objectives include determining requirements, identifying information sources, and designing a standardization process. The CRISP-DM methodology will guide the understanding of the process and data preparation, followed by modeling, evaluation, and deployment of solutions to ensure the coherence of legal data. It is evident that through the data standardization process, duplicate records are eliminated, providing a consistent format for both the names of representatives and dates.

Keywords: ADRES, data standardization, CRISP-DM, KNIME, efficiency, decision-making.

Contenido

RESUMEN ..... **V**

INTRODUCCIÓN **11**

ANTECEDENTES Y JUSTIFICACIÓN 12
FORMULACIÓN DEL PROBLEMA 13
OBJETIVO GENERAL 13
Objetivos Específicos *13*
ALCANCES Y LIMITACIONES 14

1. MARCO TEÓRICO **14**

2. DISEÑO METODOLÓGICO **17**

2.1 TIPO Y MÉTODO DE INVESTIGACIÓN 17

3. RESULTADOS **19**

4. DISCUSIÓN **29**

5. CONCLUSIONES Y RECOMENDACIONES **30**

5.1 CONCLUSIONES 30
5.2 RECOMENDACIONES 31

BIBLIOGRAFÍA **33**

Introducción

La Administradora de los Recursos del Sistema General de Seguridad Social en Salud (ADRES), establecida en virtud del artículo 66 de la Ley 1753 de 2015, desempeña un papel fundamental en garantizar la adecuada gestión de recursos, controles, pagos, giros, verificaciones y la administración de la Base de Datos Única de Afiliados (BDUA) en el sistema de salud de Colombia. Además, esta Entidad administra diversas fuentes de financiación del sistema, mientras lucha contra los desafíos asociados a los mecanismos de fraude. Debido a la gran cantidad de información que ADRES maneja, surge la necesidad de implementar una solución que permita la estandarización de su base de datos. Esta se enfocará en abordar un problema específico y será diseñada de manera que pueda adaptarse a las necesidades cambiantes de la organización, contribuyendo a mejorar la productividad y eficiencia en las tareas diarias.

La estandarización de las bases de datos es fundamental en la gestión de sistemas complejos, como el de salud en Colombia, ya que permite establecer un marco coherente y uniforme para la recopilación y el análisis de la información. En palabras de expertos en gestión de datos, como lo indica Patricio Araneda García quien destaca en su libro " Base de Datos - El camino de los datos a la información ", El valor de la información está basado en la calidad y eficiencia de los datos que la conforman (García, 2022). Asimismo, la adopción de estándares contribuye a la eficiencia operativa y a la interoperabilidad, aspectos cruciales en entidades como ADRES, donde la precisión y cohesión de los datos son esenciales para garantizar la transparencia y eficacia en la administración de recursos del sistema de salud.

En este contexto, se propone lograr una comprensión más profunda de los datos a través de un proceso de transformación que abarca desde la captura inicial del dato hasta la solución de un problema específico. Este proceso incluye la extracción, recopilación, procesamiento, almacenamiento, interpretación y visualización de la información, con el objetivo de generar datos de alto valor agregado para ADRES.

Antecedentes y Justificación

La Administradora de los Recursos del Sistema General de Seguridad Social en Salud, conocida como ADRES, desempeña un papel fundamental en el contexto de la salud en Colombia. Fundada en 2017, ADRES ha sido la entidad encargada de administrar los recursos destinados a la financiación de los servicios de salud y el aseguramiento de la población en el sistema de seguridad social en salud.

En su misión de garantizar el acceso y la calidad de los servicios de salud para todos los ciudadanos, ADRES se ha enfrentado a numerosos desafíos a lo largo de los años. Uno de los desafíos más destacados ha sido la gestión eficiente de la información legal y administrativa relacionada con su funcionamiento.

La entidad ADRES maneja una cantidad sustancial de datos legales, normativos y administrativos relacionados con los procesos de afiliación, recaudo y distribución de recursos en el sistema de salud. Estos datos son vitales para la toma de decisiones estratégicas y el cumplimiento normativo dentro de la organización. Sin embargo, a lo largo del tiempo, ha surgido un problema crítico: la falta de estandarización de bases de datos.

La falta de estandarización de bases de datos ha dado lugar a múltiples inconvenientes (Edna Sastoque, 2020), como la dificultad para acceder a información relevante de manera oportuna, la falta de coherencia en los informes legales y administrativos, y un aumento en el riesgo de errores y omisiones que pueden tener consecuencias significativas en términos de cumplimiento normativo y eficiencia operativa.

En este contexto, surge la necesidad de abordar este problema de manera integral y proactiva. Este proyecto se enfoca en comprender a fondo cómo la falta de estandarización de bases de datos afecta la operación de ADRES y, lo que es aún más importante, en proponer soluciones efectivas para mejorar la gestión de información legal. El objetivo final es llevar a ADRES a un estado óptimo en términos de toma de decisiones y gestión de la información, lo que beneficiará tanto a la organización como a la población que depende de sus servicios de salud.

A través de este proyecto, se buscará analizar críticamente la situación actual, identificar las áreas clave de mejora y diseñar estrategias que permitan a ADRES optimizar la gestión de su información legal, asegurando así, un impacto positivo en su funcionamiento y en la calidad de los servicios de salud que brinda a la comunidad.

Formulación del Problema

¿Cuál es el impacto de la falta de estandarización de bases de datos en el área de asesoría jurídica de ADRES y cuáles son las posibles soluciones para optimizar la gestión de información legal, mejorando la toma de decisiones y la eficiencia de la entidad?

Objetivo General

Diseñar una solución mediante un proceso en KNIME que permita Estandarizar las fuentes de información del proceso Reparto en el área de asesoría jurídica de ADRES.

Objetivos Específicos

- Determinar los requerimientos y necesidades del proceso de reparto de la entidad ADRES.
- Identificar las distintas fuentes de información que intervienen en el proceso de reparto de ADRES.
- Crear un proceso en la herramienta KNIME para estandarizar y unificar las fuentes de información del proceso de reparto de la entidad ADRES.

Alcances y Limitaciones

Este proyecto se realiza para generar un diagnóstico que nos permita identificar las falencias presentadas en la Administradora de los Recursos del Sistema General de Seguridad Social en Salud (ADRES), debido a la falta de estandarización en las bases de datos específicamente su información legal, generando márgenes de error significativos en la prestación de servicios, generación de reportes, toma de decisiones y otros, que puedan desencadenar inconvenientes de diferentes tipos.

1. Marco teórico

La ADRES es una Entidad autónoma del Ministerio de Salud y Protección Social de Colombia. Su función principal es administrar los recursos del Sistema General de Seguridad Social en Salud, asegurando su uso eficiente y previniendo irregularidades financieras (ADRES, 2021). En el contexto actual, la gestión eficiente de la información es vital para la toma de decisiones en las organizaciones. La oficina Asesora Jurídica de ADRES desempeña un papel fundamental en la distribución de los procesos judiciales y en el sistema de defensa de la entidad. La estandarización de las fuentes de información se convierte en un requisito esencial para lograr una gestión efectiva de este proceso.

La ADRES opera con autonomía administrativa y financiera. Su propósito es gestionar eficientemente los recursos del Sistema General de Seguridad Social en Salud para garantizar atención médica de calidad y transparencia en la administración de fondos de salud en el país. El proceso de reparto comienza con la notificación de demandas recibidas del Juzgado, registrándose a través del Procedimiento de Radicación Documental. Tras la radicación, la

solicitud se registra en la base de reparto, dando inicio al proceso de asignación a los abogados de la Oficina Asesora Jurídica (González, 2021).

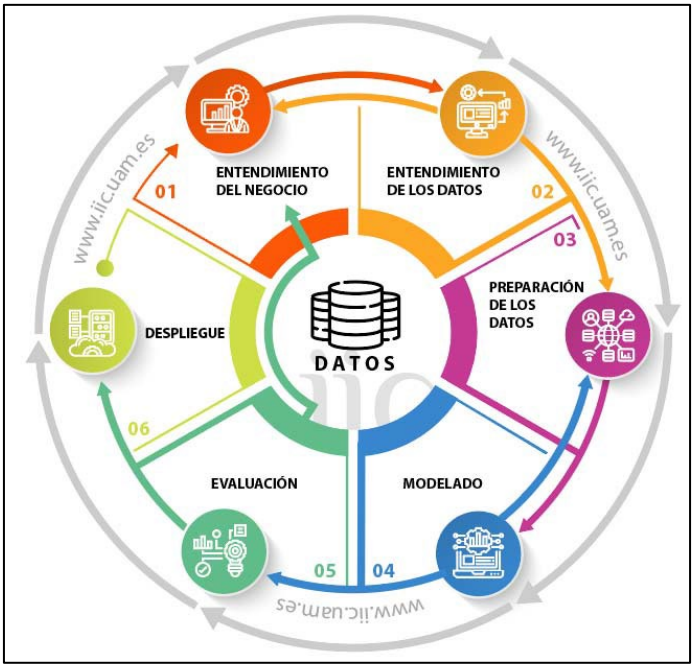
La estandarización de datos se revela como un requisito esencial en este contexto. Este proceso, también conocido como normalización, busca ajustar características para que los datos se asemejen a un modelo común, facilitando su tratamiento y acceso. Este proceso garantiza uniformidad, consistencia y formato en los datos de origen, iniciando con la interpretación de datos y reformateándolos según las necesidades organizativas. Es fundamental para asegurar la calidad y utilidad de la información, contribuyendo a una toma de decisiones más precisa y eficiente en la organización (AYUWARE, 2023) (IBM, 2021).

En la búsqueda de optimizar la gestión de la información registrada en el proceso de reparto, se propone la estandarización de datos utilizando la herramienta KNIME, plataforma open source para analítica de Business Intelligence, Machine Learning y ETL mediante un simple proceso de drag and drop (arrastrar y soltar). KNIME proporciona una interfaz gráfica de usuario donde se pueden crear flujos de trabajo rápidamente por individuos sin demasiada experiencia técnica para poder analizar los datos (TodoBI, 2020). Es una herramienta gráfica e interactiva que funciona por medio de nodos. Con KNIME se podrá realizar la construcción de un flujo de trabajo para la gestión más eficiente de la información registrada en el proceso de reparto, donde se implementarán diferentes nodos que faciliten la limpieza manual que se maneja hoy en día con la base de datos.

La adopción del marco CRISP-DM (Cross-Industry Standard Process for Data Mining) proporciona una estructura integral para este proyecto de mejora. Desde la comprensión de los objetivos y la relación con metas empresariales hasta la implementación de soluciones. CRISP-DM es un marco ampliamente utilizado para la minería de datos, compuesto por seis fases. Comienza con la "Comprensión del Negocio", donde se definen objetivos y se relacionan con metas empresariales. La "Comprensión de los Datos" implica obtener conocimiento inicial y abordar problemas de calidad. En la etapa de "Preparación de los Datos", se limpian y transforman para el análisis. La fase de "Modelado" utiliza técnicas como aprendizaje automático, que se evalúan en la siguiente etapa "Evaluación”, en esta etapa, se determina la

calidad de los modelos y se ajustan según sea necesario. La fase final, "Despliegue", implica implementar los resultados en el entorno empresarial. CRISP-DM proporciona una estructura completa para proyectos de minería de datos, desde la comprensión de objetivos hasta la implementación de soluciones (JOURNAL, 2000).

Imagen 1. Esquema del ciclo CRISP-DM estándar. (instituto de ingenieria del conocimiento, 2023)

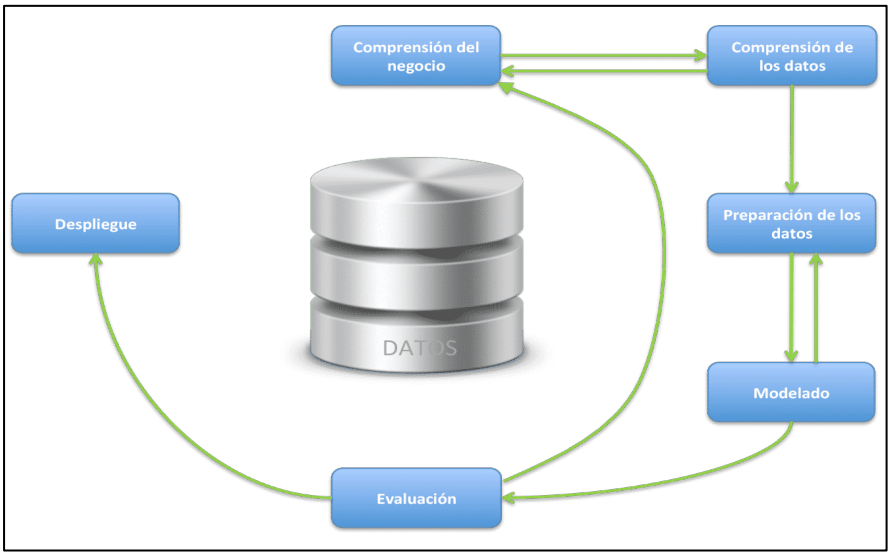


2. Diseño Metodológico

2.1 Tipo y método de investigación

Para este proyecto se utilizará una de las principales metodologías de minería de datos denominada CRISP-DM, Este diseño metodológico proporciona una estructura sistemática que permitirá guiar la investigación siguiendo el enfoque de CRISP-DM. A medida que se avance en cada fase, se documentará cada decisión, hallazgos y resultados para garantizar la trazabilidad y la comunicación efectiva a lo largo del proyecto, la imagen 2 nos muestra las etapas principales de esta importante metodología:

Imagen 2. Ciclo de vida de minería de datos, Tomada de (IBM, 2021)



A continuación, la tabla 1 contiene los tres objetivos específicos del proyecto, cada uno acompañado de sus respectivas etapas y actividades que nos permitirá visualizar de manera clara los resultados, el resumen del proceso realizado y las recomendaciones generadas.

Tabla 1. Objetivos, Etapas, Actividades y Resultados.

Objetivo Específico	**Etapa**	**Actividad**	**Resultados esperados**
Determinar los requerimientos y necesidades del proceso de reparto de la entidad ADRES.	1. Entendimiento del Negocio.	A1. Contextualización del proceso de reparto.	Resumen del proceso de reparto que permita su entendimiento.
Identificar las distintas fuentes de información que intervienen en el proceso de reparto de ADRES.	2. Entendimiento de los Datos. 3. Preparación de los Datos.	A2. Contextualización de las fuentes de datos y su problemática. A3. Diagnóstico de los datos y selección del conjunto de datos (Dataset) con el cuál se va a trabajar.	Listado de los campos o columnas de las fuentes de datos y sus principales problemas.
Crear un proceso en la herramienta KNIME para estandarizar y unificar las fuentes de información del proceso de reparto de la entidad ADRES.	4. Modelamiento. 5. Evaluación. 6. Implementación.	A4. Elaboración del proceso (stream) en la herramienta KNIME que permita Limpiar (Data Cleaning) y estandarizar las fuentes de datos. A5. Evaluación de resultados. A6. Resumen de resultados obtenidos y recomendaciones.	Proceso (stream) en la herramienta KNIME. Dataset resultado del proceso de Data Cleaning. Resumen del Proceso realizado y recomendaciones.

3. Resultados

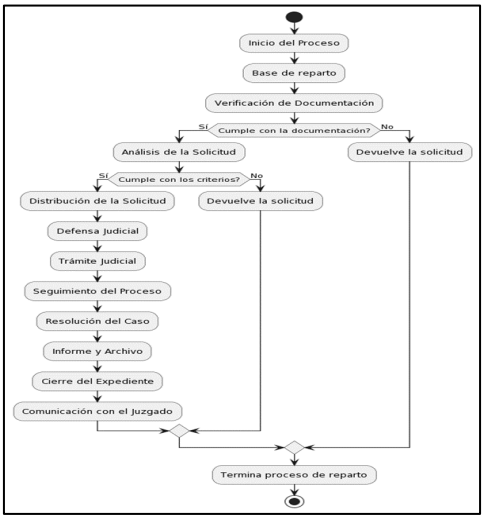
En la Tabla número 2 y la Imagen 3, se presenta un resumen detallado de las fases clave del proceso de Reparto en la Oficina Asesora Jurídica, ofreciendo una visión cronológica y estructurada de las actividades involucradas. En el proceso, se inicia con la recepción de una solicitud por parte de un juzgado o entidad judicial, marcando el inicio formal y se termina el proceso cuando se le informa al juzgado o entidad judicial sobre la resolución del caso y se toman las medidas necesarias.

Tabla 2. Resumen proceso de reparto

N°	Actividad	Descripción
1	**Inicio de proceso**	El proceso de reparto judicial comienza cuando se recibe una solicitud de un juzgado o entidad judicial.
2	**Base de reparto**	Se diligencian los datos del proceso en la base de reparto.
3	**Verificación de documentos**	Se verifica la documentación recibida para asegurarse de que esté completa y cumpla con los requisitos legales. Si cumple con la documentación continua a la actividad 4, Si No cumple se devuelve la solicitud.
4	**Análisis de documentos**	Se realiza un análisis inicial de la solicitud para determinar si se ajusta a los criterios de reparto judicial.
5	**Distribución de la solicitud**	Si la solicitud cumple con los criterios, se distribuye a un abogado o representante legal interno.
6	**Defensa judicial**	El abogado o representante legal designado asume la representación de ADRES en el proceso judicial correspondiente.

7	**Trámite judicial**	Se inicia el proceso judicial, incluyendo la presentación de escritos, comparecencias, pruebas y demás actuaciones necesarias.
8	**Seguimiento del proceso**	El equipo legal de ADRES hace un seguimiento constante del proceso judicial para garantizar una defensa efectiva.
9	**Resolución del caso**	El proceso judicial culmina con una resolución, que puede ser una sentencia judicial, un acuerdo, una conciliación o una decisión administrativa.
10	**Informe y archivo**	Una vez resuelto el caso, se genera un informe interno que registra los detalles del proceso y su resultado.
11	**Cierre del expediente**	El expediente judicial se archiva de manera adecuada y se cierra.
12	**Comunicación con el juzgado**	Se informa al juzgado o entidad judicial sobre la resolución del caso y se toman las medidas necesarias.
13	**Termina proceso**	Termina proceso de reparto.

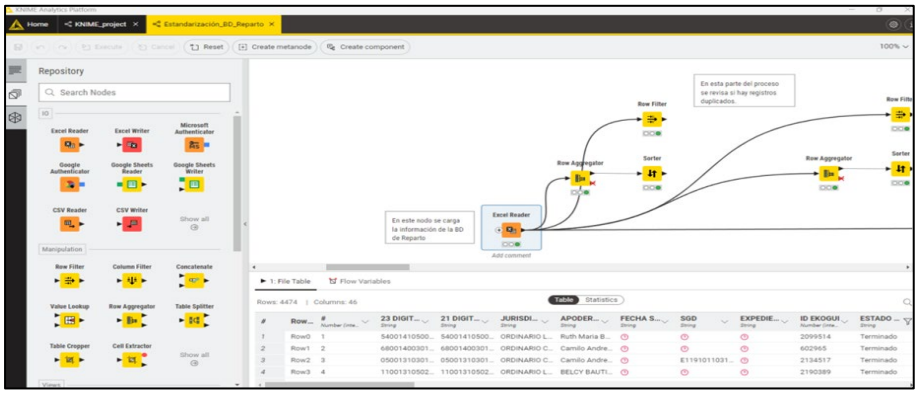
Imagen 3. Flujograma proceso de reparto



PROCESO DE ESTANDARIZACIÓN EN KNIME BD Reparto

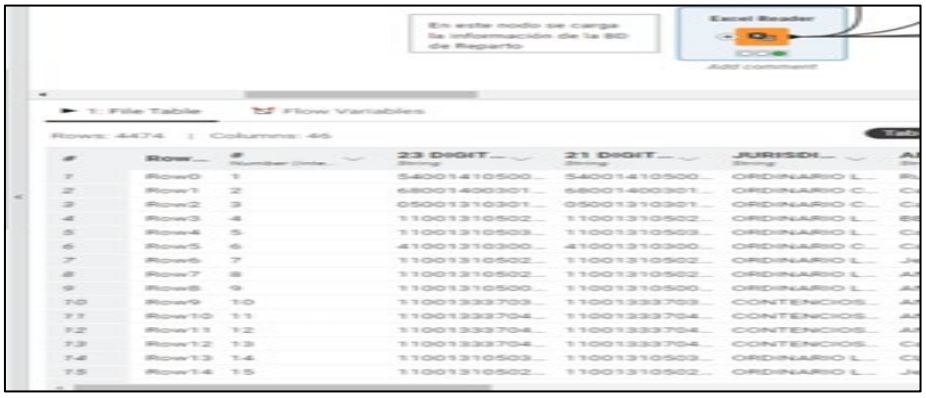
En la imagen 4 vemos la pantalla de inicio del programa KNIME, en la barra superior color oscuro aparecen las distintas pestañas con los procesos (workflows) que tenemos abiertos, a la izquierda vemos el menú con los principales nodos (Iconos), en la parte derecha encontramos el lienzo en donde construimos el flujo del proceso de estandarización con ayuda de los nodos disponibles y debajo vemos una vista previa de los nodos ejecutados, para este caso vemos como se cargó satisfactoriamente nuestra base de reparto para someterla al proceso de estandarización.

Imagen 4. Pantalla inicio herramienta KNIME.



- **Carga de la Base:** Como lo indica la imagen 5, utilizando el nodo Excel Reader se carga la información de la base de Reparto para ser utilizada en el proceso de estandarización, es importante resaltar que la base original no se ve afectada por ninguno de los procedimientos que vamos a ejecutar:

Imagen 5. Cargue base de reparto.



- **Revisión de Duplicados:** En esta parte del proceso nos concentramos en validar si existen registros duplicados haciendo una agrupación por el campo 23 DIGITOS como se observa en la imagen 6:

Imagen 6. Revisión duplicidad de registros.

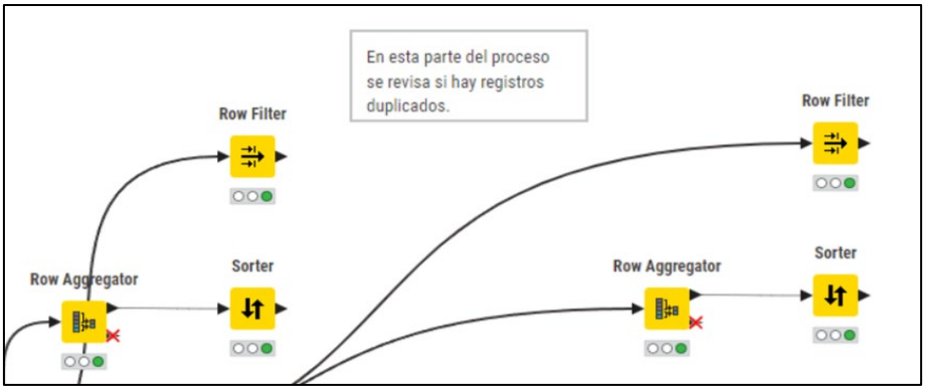


Imagen 7. Visualización registros duplicados.

#	Row...	23 DIGITOS String	OCCURRENCE_COUNT Number (long)
9	Row...	11001310502220160038900	3
10	Row...	1100131050242015002070O	3
11	Row...	11001310502620180033300	3
12	Row...	11001310503120180004500	3
13	Row...	11001310503620210013200	3
14	Row...	11001310503920190016400	3
15	Row...	11001333704020210017900	3
16	Row...	76001310500320160024300	3
17	Row20	05001233300202200567O0	2
18	Row62	05001310500620140158300	2
19	Row65	05001310500620180025600	2
20	Row69	05001310500720120104400	2

En la imagen 7, se observa una tabla con los casos de registros que se encuentran duplicados e incluso triplicados. En el siguiente paso vamos a eliminarlos.

- **Eliminación de Registros Duplicados:** En esta parte del proceso (ver Imagen 8 y 9) utilizando el nodo Duplicate Row Filter dejamos únicamente un registro por código en el campo 23 DIGITOS eliminando en el proceso los registros repetidos. Previamente se revisó el estado de esos registros duplicados siendo viable su eliminación:

Imagen 8. Visualización Nodo Duplicate Row Filter para eliminación registros duplicados.

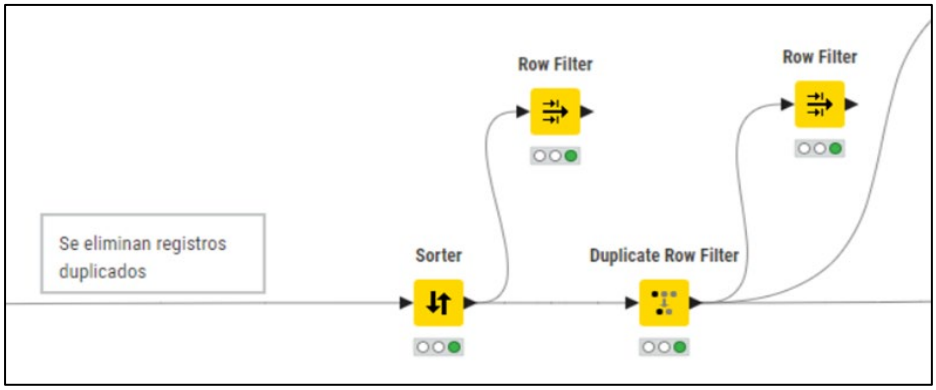
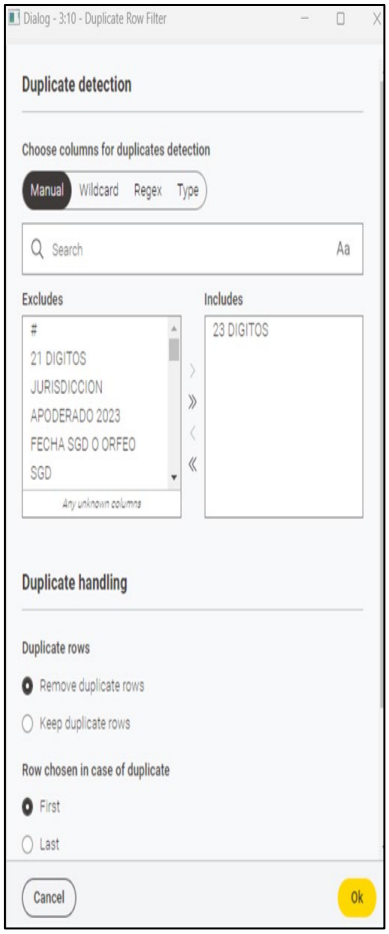


Imagen 9. Configuración registros por código.



- **Revisión y estandarización de Nombres**: Por ejemplo, en el campo APODERADO se quitaron espacios al inicio y al final de cada nombre, se dejó todo en mayúsculas y se estandarizaron los nombres de personas que a pesar de tratarse del mismo individuo estaban escritos de forma diferente (ver imágenes de 10 a 12):

Imagen 10. Revisión campo APODERADO.

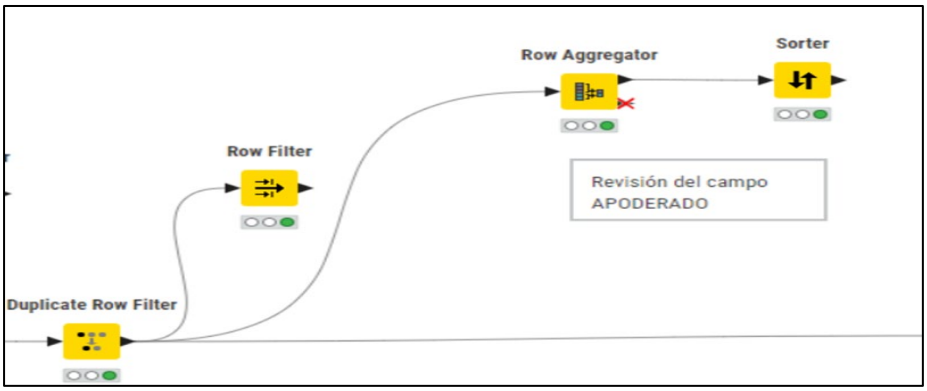


Imagen 11. Visualización estado datos APODERADO.

Rows: 44 | Columns: 2 — Table | Statistics

#	Row...	APODERADO 2023 (String)	OCCURRENCE_COUNT (Number (long))
9	Row8	ASTRID YUBEYDY VERA OQUENDO	210
10	Row9	BELCY BAUTISTA F	12
11	Row10	BELCY BAUTISTA FONSECA	164
12	Row11	Belcy Bautista Fonseca	23
13	Row12	CLAUDIA PAOLA PEREZ SUA	135

Imagen 12. Nodos utilizados en la estandarización de nombres.

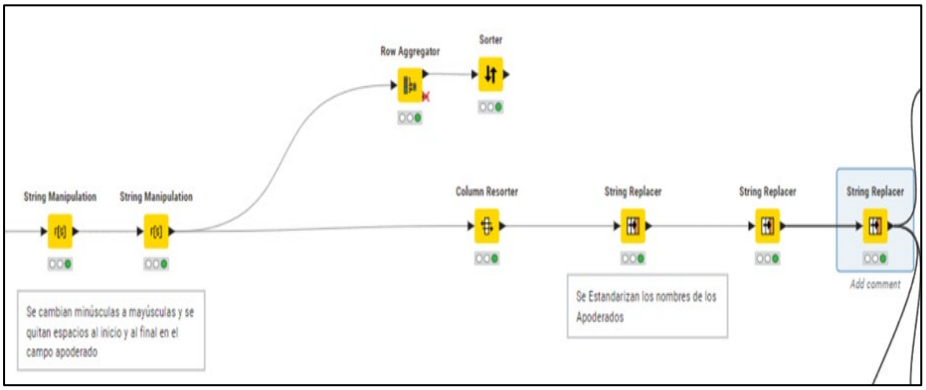


Imagen 13. Datos estandarizados.

Rows: 35 | Columns: 2 Table Statistics

#	Row...	APODERADO 2023_Dep String	OCCURRENCE_COUNT Number (long)
1	Row0	ANA C JIMENEZ A M	33
2	Row1	ANA CAROLINA JIMENEZ ACOSTA MADIEDO	193
3	Row2	ANDRES ZAHIR CARRILLO TRUJILLO	184
4	Row3	ANGIE KATERINE PINEDA RINCON	236
5	Row4	ANGIE LORENA LOPEZ PINILLA	1
6	Row5	ASTRID YUBEYDY VERA OQUENDO	210
7	Row6	BELCY BAUTISTA FONSECA	218
8	Row7	CAMILO ANDRES MOLANO	179
9	Row8	CLAUDIA PAOLA PEREZ SUA	135
10	Row9	CRISTIAN DAVID PAEZ PAEZ	127
11	Row10	DIEGO ALFONSO LEAL GONZALEZ	1

En la imagen 13 vemos como los nombres quedaron debidamente estandarizados.

- **Revisión y estandarización de fechas:** En esta parte del proceso se valida el formato de fecha, por ejemplo, el campo FECHA SGD O ORFEO tiene distintos formatos, para los registros del año 2019 algunos tienen el mes con nombres, en otros casos con números y para el caso del día algunos del 1 al 9 tienen un cero antes, pero otros no (ver imágenes 14 y 15):

Imagen 14. Revisión fechas de datos.

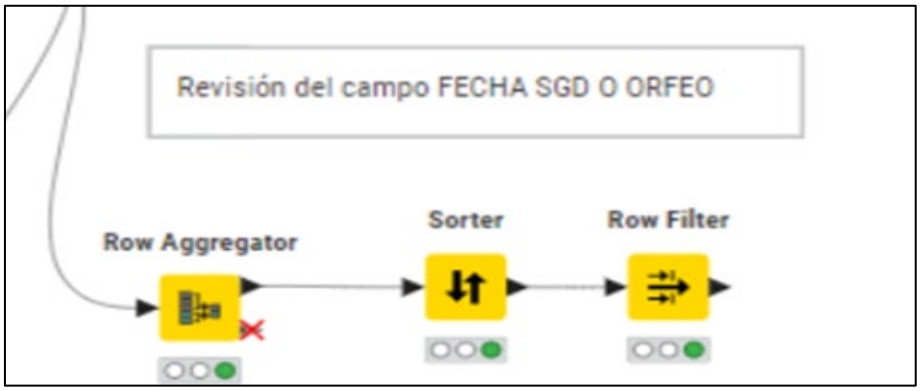


Imagen 15. Diferencias formato fecha.

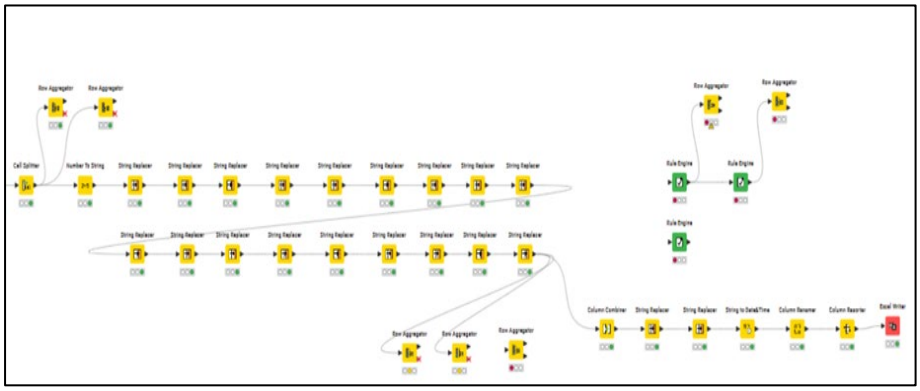
Rows: 877 | Columns: 2 — Table | Statistics

#	Row...	FECHA SGD O ORFEO String	OCCURRENCE_COUNT Number (long)
29	Row...	2023-08-04	1
30	Row...	2023-08-03	3
31	Row...	2023-08-02	2
32	Row...	2023-08-01	1
33	Row...	2023-07-31	1
34	Row...	2023-07-28	2
35	Row...	2023-07-27	2

Rows: 201 | Columns: 2 — Table | Statistics

#	Row...	FECHA SGD O ORFEO String	OCCURRENCE_COUNT Number (long)
85	Row...	2019-Febr-15	2
86	Row...	2019-Febr-14	1
87	Row...	2019-Febr-13	1
88	Row...	2019-Febr-8	1
89	Row...	2019-Febr-6	45
90	Row56	2019-12-30	1
91	Row54	2019-12-20	1

Imagen 16. Visualización general estandarización campo fecha.



En la imagen 16 vemos todos los pasos necesarios para estandarizar el campo FECHA SGD O ORFEO

- Se hace el tratamiento descrito en los pasos anteriores para cada una de las variables según sea el caso y finalmente los resultados se exportan en un archivo Excel (ver imagen 17):

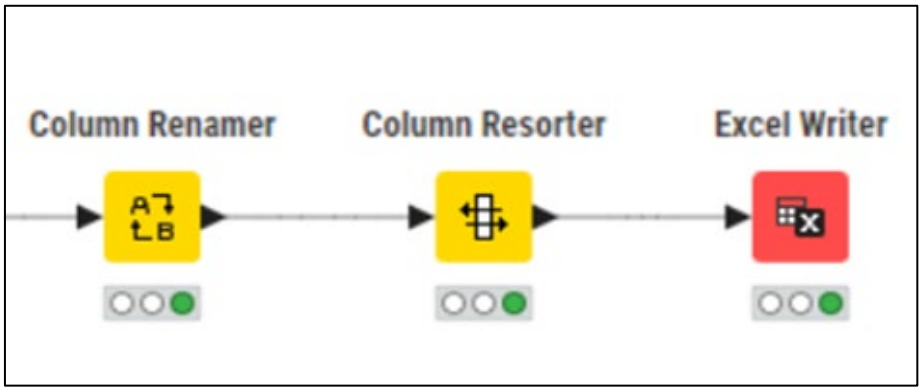


Imagen 17. Exportación a un archivo Excel.

La base estandarizada contiene campos adicionales con el sufijo _Dep para que puedan ser revisados fácilmente (ver Imagen 18):

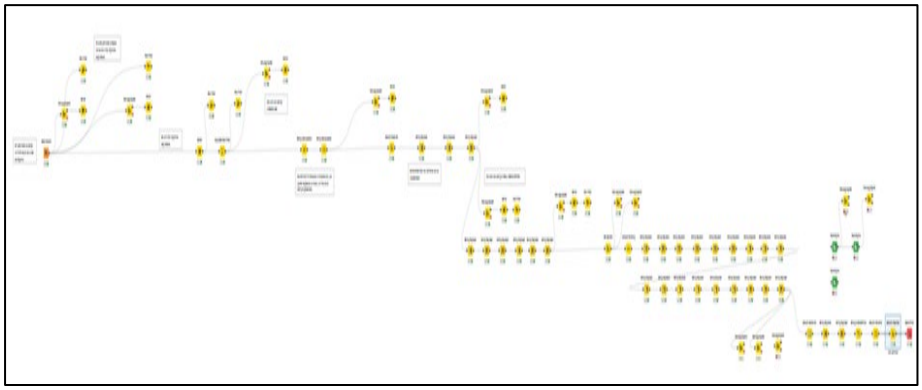
Imagen 18. Campo adicional con sufijo _Dep.

APODERADO 2023 String	APODERADO 2023_Dep String
Maria Mercedes Grimaldo Gomez	MARIA MERCEDES GRIMALDO GOMEZ
ASTRID YUBEYDY VERA OQUENDO	ASTRID YUBEYDY VERA OQUENDO
HILBANA MICHEL GALLEGO MUÑOZ	HILBANA MICHEL GALLEGO MUÑOZ
HILBANA MICHEL GALLEGO MUÑOZ	HILBANA MICHEL GALLEGO MUÑOZ
ASTRID YUBEYDY VERA OQUENDO	ASTRID YUBEYDY VERA OQUENDO
LIZSETH TATIANA CASTRILLON HERRERA	LIZSETH TATIANA CASTRILLON HERRERA
HILBANA MICHEL GALLEGO MUÑOZ	HILBANA MICHEL GALLEGO MUÑOZ
ASTRID YUBEYDY VERA OQUENDO	ASTRID YUBEYDY VERA OQUENDO

FECHA SGD O ORFEO String	FECHA SGD ORFEO_Dep Local Date
2023-08-16	2023-08-16
2019-09-27	2019-09-27
2019-02-6	2019-02-06
2021-07-30	2021-07-30
2023-03-15	2023-03-15
?	1900-01-01
2021-04-29	2021-04-29
2022-02-08	2022-02-08

A continuación, vemos la imagen 19 con el esquema general del proceso:

Imagen 19. Visualización general de todo el proceso de estandarización de datos.



4. Discusión

La ejecución del proceso de estandarización de la base de reparto muestra la necesidad de implementar políticas y controles que ayuden a mitigar los errores producto de la digitación manual durante el proceso de recolección de datos en las distintas fuentes (Edna Sastoque, 2020). Para esto se propone el uso obligatorio de un formulario que contenga validaciones y restricciones que permita únicamente el formato de texto, número o fechas, definido para cada campo o columna, así como la ejecución de un flujo en KNIME que garantice la calidad de los datos.

Encontramos que el proceso de estandarización arroja resultados positivos y deja en evidencia los principales errores a la hora de recopilar la información; duplicidad de registros, formatos de fecha inconsistentes, caracteres especiales como puntos, comas, asteriscos etc. donde no

deberían estar, nombres de personas e instituciones escritos de diferentes maneras entre otros.

La implementación del formulario plantilla y la ejecución del proceso creado en KNIME, mejorará el proceso de reparto, ya que se realiza la eliminación automática de estos errores, optimizando la digitación, generando datos más confiables y facilitando la toma de decisiones.

5. Conclusiones y recomendaciones

5.1 Conclusiones

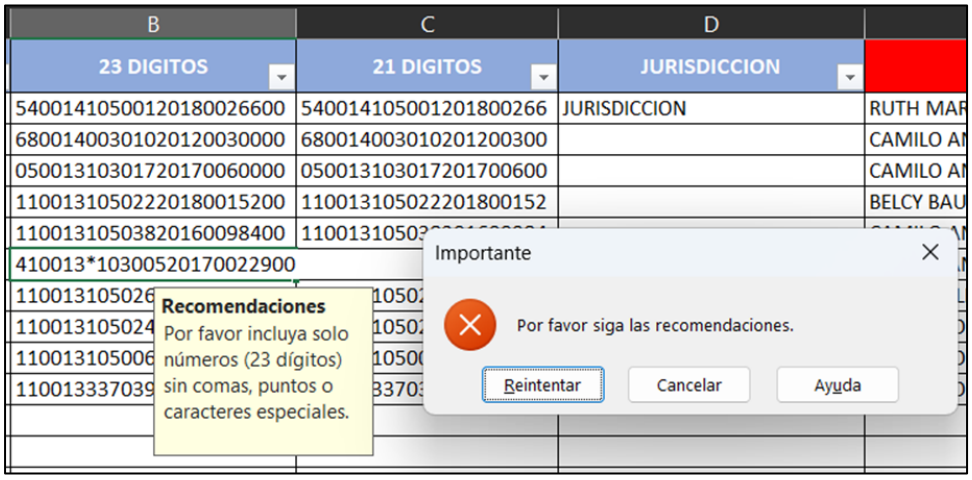
- El proceso de reparto debe mitigar drásticamente los errores presentados por la captura de información manual (Edna Sastoque, 2020), las fuentes se alimentan de datos NO estructurados como por ejemplo PDF o imágenes de documentos correspondientes a las demandas instauradas. El equipo encargado de digitar esta información y poder plasmar data no estructurara a estructurada mediante tablas de Excel, no ha encontrado una forma que permita optimizarlo.
- Es importante la implementación de políticas y controles que ayuden a mitigar los errores por digitación manual de registros, acompañados de capacitaciones al equipo de digitación, relacionada con las buenas prácticas de recolección de datos, estas deben ser implementadas en cada una de las fuentes de información que alimenta la base de Reparto.

- El proceso de estandarización, si bien corrige el problema y los errores presentados, debe estar acompañado de acciones adicionales, previas a la consolidación de la información, un claro ejemplo es el uso de formularios.
- La falta de la estandarización en las bases de datos presenta un impacto significativo en la eficiencia y toma de decisiones, dado que las diferentes fuentes de información limitan su capacidad de respuesta, debido a esto se implementa el proceso de estandarización en KNIME, esta permite interactuar con la data de forma más versátil y junto con las políticas y controles implementados, eliminan las necesidades durante el proceso de reparto.

5.2 Recomendaciones

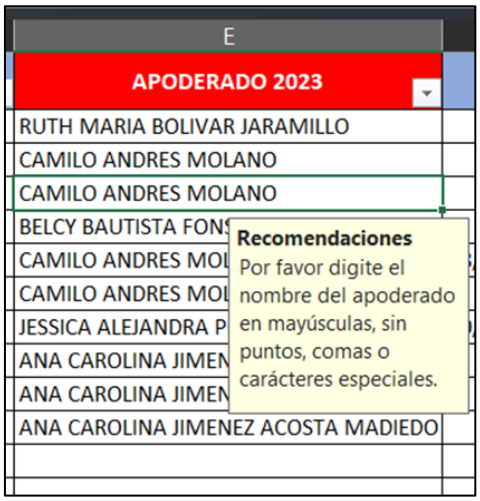
- Es importante que para la captura de información utilicen un formulario o plantilla con instrucciones, listas desplegables y las debidas protecciones para evitar que registren caracteres especiales asegurándose que todos utilicen los mismos formatos en los campos diligenciados, esto reducirá significativamente la posibilidad de cometer errores por digitación.

Imagen 20. Visualización recomendaciones campo 23 DIGITOS.



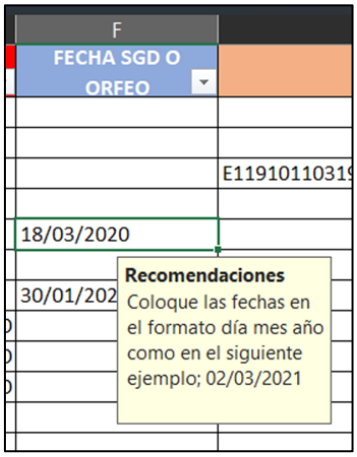
En la imagen 20 podemos ver un ejemplo del formulario sugerido con las recomendaciones para diligenciar el campo 23 DIGITOS de forma correcta y un mensaje alertando cuando se ingresa algo no válido.

Imagen 21. Visualización recomendaciones campo APODERADO 2023.



Siguiendo con los ejemplos del formulario, la imagen 21 muestra recomendaciones para el correcto diligenciamiento del campo APODERADO 2023.

Imagen 22. Visualización recomendaciones campo FECHA SGD O ORFEO.



Finalmente, la imagen 22 nos muestra un ejemplo con las recomendaciones para los campos tipo fecha.

- Se puede hacer un archivo que relacione de forma correcta las distintas entidades mencionadas en las demandas y que alimentan las listas desplegables del formulario recomendado en el punto anterior, este archivo debe mantenerse actualizado.
- Se recomienda capacitar al personal en temas relacionados con buenas prácticas de digitación, socializando la forma correcta de diligenciar la información de los campos necesarios en la base de reparto.
- Recomendamos además del uso del formulario sugerido, ejecutar el proceso de estandarización a las demás variables de la base Reparto y hacerlo periódicamente al menos 1 ves al mes.

Bibliografía

ADRES. (2021). *ADRES*. Obtenido de https://www.adres.gov.co/nuestra-entidad/acerca-de-adres

AYUWARE. (2023). *AYUWARE*. Obtenido de https://www.ayuware.es/blog/como-estandarizar-datos/

Edna Sastoque, A. C. (17 de 08 de 2020). *razonpublica*. Obtenido de razonpublica: https://razonpublica.com/los-retos-del-sistema-judicial-colombia/

González, R. A. (18 de 11 de 2021). *ADRES*. Obtenido de https://www.adres.gov.co/planeacion/procesos-y-procedimientos/Gestin%20Jurdica/Representaci%C3%B3n%20Judicial.pdf

IBM. (17 de Agosto de 2021). *Ciclo*. Obtenido de IBM: https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview

IBM. (28 de 02 de 2021). *IBM*. Obtenido de https://www.ibm.com/docs/es/iis/11.5?topic=jobs-standardizing-data

Imagine Easy Solutions. (09 de 2016). *Normas APA*. Obtenido de http://normasapa.net/

instituto de ingenieria del conocimiento. (2023). *Esquema*. Obtenido de instituto de ingenieria del conocimiento: https://www.iic.uam.es/innovacion/metodologia-crisp-dm-ciencia-de-datos/

JOURNAL. (2000). *mineria de datos*. Obtenido de https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf

Mendeley. (09 de 2016). *Mendeley*. Obtenido de https://www.mendeley.com/

Microsoft. (2023). *Microsoft*. Obtenido de https://support.microsoft.com/es-es/office/tareas-b%C3%A1sicas-en-excel-dc775dd1-fa52-430f-9c3c-d998d1735fca

Roy Rosenzweig Center for History and New Media. (09 de 2016). *Zotero*. Obtenido de https://www.zotero.org

Sáenz, J. (09 de 09 de 2016). *Wiki Sistemas Digitales*. Obtenido de http://sistdig.wikidot.com

Suárez, R. (2015). Energías renovables, Impacto y Efectividad. *Revista Científica Ingeniería y Desarrollo en Energía*, 48-53.

Thomson Reuters. (09 de 2016). *EndNote*. Obtenido de http://endnote.com/

TodoBI. (28 de 07 de 2020). *TodoBI*. Obtenido de https://todobi.com/introduccion-a-kine/