Fortalecimiento de la gestión de información del Centro Nacional de Memoria Histórica para toma de decisiones en instancias de coordinación y articulación

Diani Yeimi Celeita Barrera, Enrique Eduardo Orozco Pacheco, Jean Nicolae Gutiérrez Alfonso, John Jairo Rincón Chingate y Rodrigo Antonio Rojas Tolosa

Especialización en Big Data, Fundación Universitaria Compensar

Propuesta de consultoría

Wilson Hernando Soto Urrea

27 de Noviembre de 2023

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

Fortalecimiento de la gestión de información del Centro Nacional de Memoria Histórica para toma de decisiones en instancias de coordinación y articulación

Diani Yeimi Celeita Barrera, Enrique Eduardo Orozco Pacheco, Jean Nicolae Gutiérrez Alfonso, John Jairo Rincón Chingate y Rodrigo Antonio Rojas Tolosa

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

Especialista En Big Data

Director:

Ingeniero Wilson Hernando Soto Urrea

Fundación Universitaria Compensar

Facultad de Ingeniería, Especialización en Big Data

Bogotá, Colombia

2023

Resumen

En este proyecto de consultoría, se presentará una propuesta de mejoramiento de datos basada en un diseño principalmente de investigación cuantitativo no experimental de tipo documental al Centro Nacional de Memoria Histórica (CNMH) en la ciudad de Bogotá D.C. Esta propuesta buscará mejorar procedimientos para la gestión de información en la entidad, para que influya positivamente en la medición de las metas comprometidas en instancias de coordinación y articulación relacionadas con el deber de memoria del Estado y el esclarecimiento de la verdad, como contribución a la reparación a las víctimas del conflicto armado interno en Colombia. Para ello, planteamos la formulación de un plan de mejoramiento y gestión de la información desde la metodología de Minería de datos – Modelo CRISP-DM, que le permitirá a la Entidad, tener plena identificación de sus fuentes de datos y procesos para la gestión de metas e indicadores, así como una propuesta para la adopción de herramientas de minería de datos que le facilitarán al CNMH contar con un control adecuado y automatizado de sus indicadores.

Palabras clave: Instancia de coordinación y articulación, Minería de datos, Modelo CRISP-DM, Gestión de información

Abstract

In this consultancy project, a data improvement proposal based on a mainly quantitative research design, non-experimental, of documentary type will be presented to the National Center of Historical Memory (CNMH) in of Bogotá D.C. This proposal will seek to improve procedures for information management in this Entity to positively influence the measurement of the goals committed in coordination and articulation instances related to the duty of the state to preserve memory and the clarification of the truth, as a contribution to the reparation of the victims of the internal armed conflict in Colombia. For this purpose, we propose the formulation of an improvement and information management plan from the methodology of Data Mining - CRISP-DM Model, which will allow the entity to have full identification of its data sources and processes for the management of goals and indicators, as well as a proposal for the adoption of data mining tools that will facilitate the CNMH to have an adequate and automated control of its indicators.

Key words: Coordination and articulation body, Data mining, CRISP-DM model, Information management.

Contenido

PÁG.

RESUMEN ..... **V**

INTRODUCCIÓN **11**

ANTECEDENTES Y JUSTIFICACIÓN 11
FORMULACIÓN DEL PROBLEMA 12
OBJETIVO GENERAL 12
Objetivos Específicos *13*
ALCANCES Y LIMITACIONES 13

1. MARCO TEÓRICO **15**

2. DISEÑO METODOLÓGICO **17**

3. RESULTADOS ESPERADOS **23**

4. CONCLUSIONES Y RECOMENDACIONES **24**

4.1 CONCLUSIONES 24
4.2 RECOMENDACIONES 25

BIBLIOGRAFÍA **26**

Introducción

Para fortalecer el deber de memoria del Estado y la contribución al esclarecimiento de la verdad sobre las violaciones ocurridas en el marco del conflicto armado y como parte de la reparación integral a las víctimas (Centro Nacional de Memoria Histórica, 2023), esta propuesta pretende la creación de un plan de acción para mejorar los procedimientos utilizados en la gestión de los datos y posterior visualización de los mismos, otorgando significado y utilidad como información para la toma de decisiones, a través de la implementación de una metodología de minería de datos. Esto, en búsqueda de que el CNMH, fortalezca el modelo de gestión de conocimiento e innovación actual y brinde un aporte importante en la formulación de una política pública de memoria y verdad.

Este plan de trabajo se centrará en los procesos de gestión de información relacionados con metas asociadas con actividades, productos y resultados que el CNMH tiene a cargo en instancias de articulación y coordinación.

Antecedentes y Justificación

En virtud del deber de memoria del Estado, consagrado en la Ley 1448 de 2011 (2016) y las normas reglamentarias, el CNMH, viene trabajando en la garantía de los derechos de las víctimas del conflicto armado interno a la verdad, la justicia, la reparación y la no repetición (Centro Nacional de Memoria Histórica, 2020).

El CNMH se encuentra ante el reto de reorientar su marco estratégico, atendiendo a cambios importantes en su entorno, considerando la discusión alrededor de conceptos fundamentales de su misión, visión y objetivos estratégicos (Centro Nacional de Memoria Histórica, 2020), en términos de su contribución a la reconstrucción de

memoria histórica y al esclarecimiento de la verdad, así como a la justificación de su pervivencia más allá del marco transicional en el que fue creado.

Repensar el marco estratégico de la entidad ha implicado considerar su articulación y coordinación con otras entidades y actores, así como su modelo de operación interna. Para ello se encuentra desarrollando iniciativas como una consultoría para el diseño de su Modelo de Gestión de Conocimiento e Innovación y otra consultoría para la elaboración de los capítulos de un documento CONPES (Agencia Cooperación Alemana. 2023) de formulación de política de memoria y verdad. Así mismo, se encuentra adportas de la formulación de nuevos proyectos de inversión a partir de 2024 (Centro Nacional de Memoria Histórica, 2023). Estas iniciativas justifican revisar y considerar qué cambios se requieren en la gestión de datos e información relacionados con metas de gestión, productos y resultados asociados a instancias de coordinación y articulación en las cuales se encuentra involucrado el CNMH.

Por lo anterior y con el fin de poder ayudar al CNMH a lograr sus objetivos, para esta necesidad es pertinente la implementación de una metodología de minería de datos que permita realizar un entendimiento de la propuesta de valor de la entidad, el entendimiento y preparación de datos para facilitar el modelado de una solución que posterior a ser evaluada, permita con su despliegue un control adecuado y automatizado de sus indicadores.

Formulación del Problema

¿Cuáles son los procesos que debería implementar el CNMH que le permitan mejorar sus prácticas y gestión de información frente a las metas e indicadores en las instancias de articulación y coordinación en los que participa?

Objetivo General

Mejorar la gestión de la información en el Centro Nacional de Memoria Histórica para la medición de metas comprometidas en instancias de coordinación y articulación.

Objetivos Específicos

1. Identificar las fuentes de datos y procesos involucrados en la gestión de metas relacionadas con actividades, productos y resultados del CNMH en instancias de coordinación y articulación.

2. Proponer técnicas y herramientas de transformación de datos que permitan ejecutar las diferentes metodologías de minería de datos requeridas en el modelo a plantear.

3. Proponer un plan de acción para el CNMH a partir de los resultados obtenidos de la metodología de minería de datos seleccionada que contemple su evaluación, aceptación y posterior despliegue en la entidad.

Alcances y Limitaciones

La consultoría se concentra en analizar las oportunidades de mejora observadas en la gestión de datos de metas e indicadores de la entidad, específicamente en relación con la toma de decisiones. Para brindar claridad y contexto a la delimitación del problema, se establecen las siguientes consideraciones:

Temporal: El enfoque se dará en los procesos para la gestión de datos de metas e indicadores en el CNMH para un rango temporal correspondiente a la vigencia del actual Plan Nacional de Desarrollo 2022 – 2026 (Departamento Nacional de Planeación, 2022), es decir, desde 2022 hasta el último mes vencido al momento del inicio de la consultoría. Además, en términos de oportunidad, se pretende ejecutar esta consultoría en el primer semestre 2024, entendiendo que en este periodo se formularán los nuevos proyectos de inversión y se desarrollará la segunda fase de las consultorías de gestión de conocimiento y de política pública de memoria y verdad.

Metodológico: Esta consultoría estará basada en un diseño de investigación principalmente cuantitativo no experimental de tipo documental (Hernández et al. 2018), incluyendo, entrevistas con personal del CNMH y análisis de datos disponibles en sistemas internos. Esta propuesta no contempla el desarrollo de software sino el plan

de trabajo para la ejecución de un modelo de minería de datos y los lineamientos de evaluación y despliegue.

Recurso Humano: Inicialmente, se cuenta con un equipo base interdisciplinario para la ejecución de la consultoría. Sin embargo, se requerirá el apoyo y acompañamiento del CNMH y sus funcionarios, a fin de realizar levantamiento y procesamiento de información, análisis y adopción de los resultados obtenidos y capacitación al equipo de TI y otros involucrados para las implementaciones.

Recursos financieros: Se ha identificado como potencial fuente de financiación el programa ProPaz de la Agencia de Cooperación Alemana (GIZ), dado que la propuesta es afín a los campos de acción del Programa y es la fuente actual de las consultorías de Modelo de Gestión de Conocimiento e Innovación para el CNMH y de formulación de política pública de memoria y verdad. (2022).

Limitaciones: Se debe considerar el estatus preliminar a la información (Aún en construcción) en el CNMH. También, aclarar a la entidad que esta consultoría se enfocará en los procesos y datos gestionados por la entidad. Por tanto, no involucrará gestión de fuente de datos en otras entidades.

1. Marco teórico

Metodologías de minería de datos:

Para determinar las técnicas, metodologías y modelos que permitan desarrollar esta propuesta, se considera necesario realizar un estudio de los métodos más populares para realizar minería de datos en las empresas (KDD, CRISP-DM y SEMMA) para realizar una selección consciente y precisa que permita desarrollar los objetivos propuestos.

Estos métodos, permitirán dar una guía concisa de la implementación de la minería de datos y darán al CNMH una mejor opción para el uso y manejo de la información eficiente. (Shafique et al. 2014). Para lo anterior, se generará una comparación de los modelos propuestos y así elegir el modelo de minería de datos que se desarrollará posteriormente enfocado a las necesidades de la Entidad:

Tabla 1.
Comparaciones de metodologías de minería de datos

Concepto	KDD	SEMMA	CRISP-DM
Definición	Modelo iterativo e interactivo que permite descubrir patrones y conocimiento oculto en grandes conjuntos de datos.	Permite la comprensión, organización, desarrollo y mantenimiento de proyectos de minería de datos.	Proporciona un marco uniforme y otorga directrices para la minería de datos.
Fases	1. Desarrollo y entendimiento 2. Creación de conjunto de datos objetivo 1. Desarrollo y entendimiento del tema 2. Creación de conjunto de datos objetivo 3. Limpieza y pre procesamiento de datos 4. Transformación de datos 5. Elección de la tarea de minería de datos 6. Elección del algoritmo de minería de datos 7. Ejecución del algoritmo de minería de datos 8. Interpretación de la información y patrones obtenidos 9. Conclusión y usabilidad del conocimiento encontrado	1. Muestreo de datos 2. Exploración de datos 3. Modificación de datos 4. Modelado de datos 5. Evaluación de los resultados	1. Entendimiento del negocio 2. Entendimiento de la data 3. Preparación de la data 4. Modelado 5. Evaluación 6. Despliegue
Enfoque Iterativo	Sí, las fases pueden repetirse iterativamente.	Sí, especialmente la fase de Modelado.	Sí, cada fase puede retroalimentar las anteriores.
Flexibilidad	Alta flexibilidad debido a su naturaleza general.	Menos flexible en comparación con KDD.	Flexible, pero sigue un enfoque estructurado.
Énfasis en la Exploración	Enfoca en la extracción de patrones y conocimiento.	Fuerte énfasis en la exploración de datos.	Pone énfasis en la comprensión inicial de los datos.

Nota. Comparación de las metodologías de minería de datos más usadas para el análisis de información. (Ludus, 2018 & Moine, ed al. 2011).

Knime: Es una plataforma gratuita y de código abierto, que permite realizar de forma intuitiva, la transformación y procesamiento de datos. También tiene nodos de visualización que permite al usuario tener de forma integral todos sus procesos en una sola herramienta. Cuenta con más 300 conectores a diferentes fuentes de datos y a bibliotecas, lo que permite tener una gran cobertura al momento de utilizar diferentes extensiones de archivos. (KNIME Analytics Platform, s.f.)

Python: Es un lenguaje de código abierto, sencillo de utilizar, amigable, fácil de aprender y reduce costos financieros. Se puede usar en minería de datos por su versatilidad, simplificando tiempos y esfuerzos en la construcción del código base. (Python, s.f.)

Matplotlib: Es una biblioteca creada sobre lenguaje de programación de Python, la cual permite crear visualizaciones estáticas, animadas e interactivas de la información procesada y cargada. (Matplotlib, s.f.)

Pandas: Es una biblioteca que proporciona herramientas de manipulación de datos y análisis de datos, permitiendo la manipulación de estos de forma segura y contribuyendo a la optimización del rendimiento. (Pydata, s.f.)

Numpy: Es una biblioteca que brinda soporte en objetos de matriz multidimensional, matriz enmascarada, matemáticas, operaciones estadísticas básicas, algebra lineal, permitiendo a los usuarios realizar de forma ágil cálculos matemáticos. (Numpy, s.f.)

Seaborn: Es una biblioteca basada en matplotlib destacándose por su interfaz de alto nivel en la generación de gráficos estadísticos llamativos y muy productivos a la hora de contribuir con la toma de decisiones. (Pydata. S.f.)

Colab: Es un producto de Google Research que le permite al usuario usar y ejecutar código Python de forma gratuita. Dentro de sus características permite realizar un aprendizaje automático y análisis de datos. (Google. S.f.)

Looker Studio: Es una herramienta que pertenece a Google Workspace, la cual permite interactuar con los datos, transformarlos y crear visualizaciones (tableros de control) que permiten al usuario, observar la data procesada de una forma intuitiva y dinámica. Esta interactúa con diferentes fuentes de datos, lo que permite una integración global con varias extensiones de archivos. (Google, s.f.)

2. Diseño Metodológico

Teniendo en cuenta la información por el integrante del equipo conocedor del CNMH, se pudo determinar que, para la medición de metas en instancias de articulación y coordinación, el proceso requiere de alrededor de 130 bases de datos de diferentes fuentes origen, lo que equivale a 24650 datos aproximados (Obtenidos de diferentes periodicidades) que se entregan de forma no estandarizada, en el repositorio de información Google Drive a través de carpetas en unidades compartidas y/o a través de la herramienta Excel o Google Sheets.

Dado lo anterior, la metodología CRISP-DM considera la interdependencia entre fases y la adaptación de las mismas dependiendo de la necesidad (IBM, 2021) . En este sentido, la metodología propuesta abordará las fases de entendimiento de negocio y de datos de forma interdependiente y simultánea, mientras que las demás serán secuenciales, pero iterativas. Para ello, se debe considerar:

Entendimiento del negocio:
Esta fase abordará, desde un enfoque cualitativo, los factores que, desde los fundamentos para la dirección de proyectos, se considerarán relevantes dentro de un sistema para la entrega de valor.

- Componentes de la entrega de valor: este factor se refiere al entorno interno del CNMH que permite la disposición de bienes o servicios con los cuales contribuye a la reconstrucción de la memoria histórica y al esclarecimiento de la verdad. Se abordará, de una parte, la cadena de valor de los seis proyectos de inversión de la entidad y su alineación con sus objetivos estratégicos; de otra parte, la operación del Modelo Integrado de Planeación y Gestión de la entidad, en especial la dimensión Direccionamiento estratégico y planeación, la Dimensión evaluación de resultados, el esquema operativo institucional y la Dimensión Gestión del Conocimiento y la Innovación. (Centro Nacional de Memoria Histórica, 2023)

- Flujo de información desde el direccionamiento estratégico hacia las operaciones sobre los resultados, beneficios y valor deseados, y viceversa, en donde desde las operaciones se brinda información sobre ajustes y actualizaciones requeridos para mejorar la entrega de valor. Para ello se recurrirá a la diagramación jerárquica a través del Diagrama de desglose de la organización, que ofrece una representación sobre la relación entre actividades y las unidades de la organización. Así mismo, se realizará revisión documental de informes sobre caracterización de sistemas de información y de diagnóstico de gestión de conocimiento presentados en la consultoría especializada para el diseño del Modelo de Gestión de Conocimiento y la Innovación del CNMH. (Centro Nacional de Memoria Histórica, 2023)

- El sistema de gobernanza del CNMH entendido como los elementos de supervisión, control, evaluación de valor, integración entre componentes y capacidades para toma de decisiones. Estos elementos se examinarán para cada una de las cuatro direcciones técnicas de la entidad. Para ello se realizará revisión documental del organigrama de la entidad, complementando y contrastando con entrevistas semiestructuradas a los enlaces grupo de Planeación. (Centro Nacional de Memoria Histórica, 2023)

- La vinculación la entidad a instancias de coordinación y articulación. En este caso se entenderá por instancias, de una parte, los espacios de coordinación y articulación en los que la entidad tiene a cargo metas en planes operativos de esos espacios. Así mismo, dentro del concepto de instancias se incluyen los instrumentos de política pública en los cuales el CNMH tiene acciones comprometidas. Adicionalmente, se considerarán dentro del concepto de instancias, los espacios de concertación con sujetos colectivos con quienes la entidad ha adquirido compromisos frente a acciones a desarrollar. Se abordará a través de revisión documental de los planes de acción u otros documentos programáticos que sean suministrados por el Grupo de Planeación de la entidad. (Centro Nacional de Memoria Histórica, 2020)

Entendimiento de los datos:

Con el fin de poder realizar el entendimiento de las fuentes de datos necesarias para determinar el avance de metas, se debe realizar un proceso que implica la recopilación de datos iniciales, su descripción, exploración y la verificación de su calidad. En esta etapa se comprende:

- Recopilación de datos iniciales. Dado que el CNMH gestiona sus datos e información en Google Workspace, las fuentes de datos se recopilarán en carpetas de unidad compartida de Drive, para acceso compartido por parte de los integrantes de la consultoría y del Grupo de Planeación de la entidad.
- Conocer los tipos de datos. Se elaborará un diccionario de datos para describir la etiqueta del dato; si es cualitativo o cuantitativo discreto o continuo; su escala de medición para saber si es nominal, ordinal, de intervalo o razón; el tipo de dato en términos de clasificar si es entero, flotante, cadena, booleano, etc.
- Exploración de los datos y verificación de su calidad. Para esta actividad se cargarán los datos en el aplicativo Knime, que permite una mejor integración con diferentes servicios de Google Workspace y que a su vez permitirá realizar un análisis exploratorio de datos en su estructura a través de los principales estadísticos descriptivos y representaciones gráficas, identificación de duplicados, variables innecesarias, registros incompletos, datos atípicos, entre otros.

Con el desarrollo de esta fase se pretende contar con un nivel de comprensión de los datos que permita identificar las fuentes de datos origen, sus atributos clave, la reformulación de hipótesis en dado caso que los datos encontrados no sean aplicables al modelo, los problemas de calidad de los datos, así como otras necesidades de integración o desagregación que se requieran antes de su preparación para el modelado.

Preparación de los datos:

En esta fase se tomarán las fuentes de datos recopiladas y exploradas en el punto anterior, para aplicar los procedimientos que se requieran en este paso. Esto puede

implicar operaciones de selección, limpieza, transformación o construcción de nuevos datos, integración de datos, clasificación, así como dar formato a los datos para el modelado. A continuación, se describen las actividades a desarrollar en este proceso.

- Selección de datos: Se usará la herramienta Knime, la cual es de código abierto y con una interfaz intuitiva que permite crear flujos de trabajo y nodos personalizados de acuerdo a las necesidades (KNIME Analytics Platform, s.f.), tanto para la preparación de datos, como para el modelado. Para la selección de los datos se utilizarán los nodos Google Sheets Reader, Google Sheets writer y CSV reader y CSV writer. Así mismo, el grupo de nodos Manipulación.
- Limpieza de datos. Partiendo de los resultados del análisis exploratorio de datos se procederá al tratamiento de valores perdidos y atípicos, corrección de errores de entrada, reducción de inconsistencias y eliminación de duplicados.
- Transformación de los datos. De acuerdo con las necesidades del modelado se utilizarán técnicas de codificación, estandarización y normalización,
- Construcción de nuevos datos. Si el modelado lo requiere, se crearán nuevos datos a partir de la creación de variables derivadas y la agrupación o segmentación de datos.
- Para las actividades de limpieza, transformación y construcción de datos se utilizarán los nodos de Knime de los grupos Manipulation, Convert & Replace, Transform, 20plit & combine, entre otros que se requieran de acuerdo con la necesidad (KNIME Analytics Platform, s.f.). De manera complementaria si las opciones de Knime no son suficientes, se recurrirá a la importación de librerías para Python en Colab de Google para el desarrollo de estas actividades. (Google, s.f.)

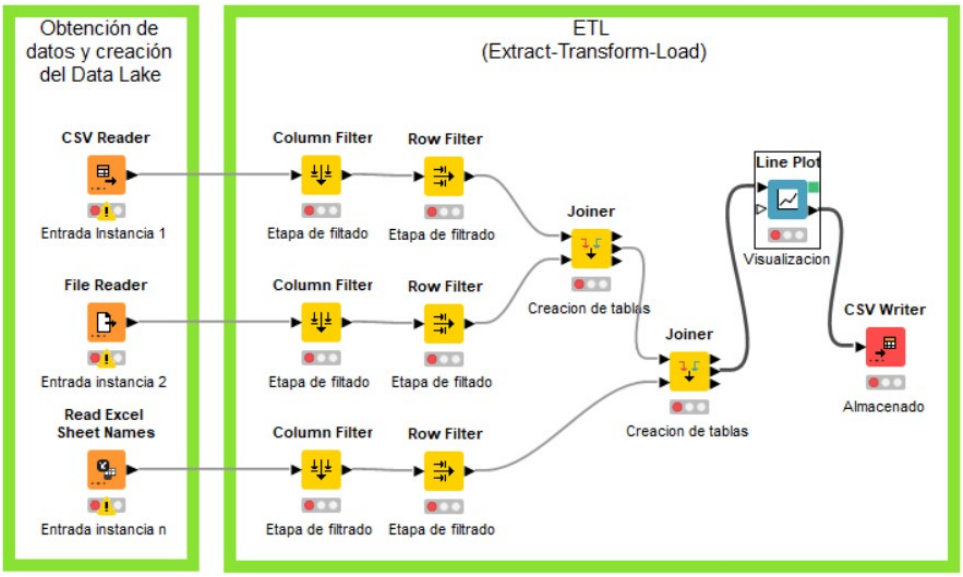
Modelado:

Para el modelado, es importante saber con qué tipos de datos cuenta el CNMH, para esto se basará en el DataLake ya estructurado anteriormente y alineado a los objetivos que se desean cumplir con relación a metas de la entidad.

Se implementará un modelado de los datos con el fin de tener esquemas o representaciones gráficas de los datos obtenidos en la fase anterior. Con esto se pretende brindar una ayuda visual al CNMH para la validación del comportamiento de estos.

Imagen 1.

Fases del modelado usando software Knime



Nota. Elaboración propia aplicativo Knime

Evaluación

Esta etapa es una de las más importantes dentro de la metodología CRISP-DM, ya que permite determinar si se cumplen con los objetivos propuestos en esta consultoría. Además, se puede establecer, si los resultados dan un índice sobre algún error especifico en alguna etapa anterior. (Israel. 2022).

Es importante recalcar que, si los resultados fueron exitosos y cumplen con los objetivos propuestos, está etapa permite realizar la validación y brindar el visto bueno de la metodología usada. Es por esto que se debe validar detalladamente los siguientes pasos:

- Validación de la implementación de las técnicas de limpieza y transformación de datos en las fases anteriores. Esto buscará generar mejoras sustanciales en la calidad general de la información que maneja el CNMH. Si los datos no son los

esperados, se realizará la alerta para indicar en que etapa se debe revisar la inconsistencia. (IBM, s.f.)

- La visualización que se generará en la plataforma KNIME desempeña un papel fundamental en la presentación de resultados. Esta herramienta integral de análisis de datos, proporciona facilidades para la preparación, transformación y análisis de datos, logrando una consolidación y estructuración eficiente de la información proveniente de diversas fuentes. Igualmente, si esta herramienta no cumple las expectativas generadas, se realizará la identificación y plan de acción según su etapa. (Haya, 2021)

Despliegue.

Una vez evaluados los resultados y con el fin de dar a conocer los nuevos conocimientos que arrojó el proceso de modelado para la implementación de las mejoras en el CNMH, se deben ejecutar las siguientes actividades:

- Se deben generar las políticas de planificación y control que permitan generar un despliegue seguro y con el alcance a toda la organización. (IBM, s.f.). Este proceso, debe incluir el resumen de los resultados, creando una planificación para el paso a producción a nivel tecnológico, el plan de difusión de los resultados y los planes de contingencia para eventuales problemas en la implementación.

- De acuerdo con la necesidad del CNMH, es importante generar un plan de control y mantenimiento posterior al desarrollo de este proyecto, que permita a la organización monitorear el efecto de los cambios a establecer. Para ello, se deben crear indicadores que permitan realizar el monitoreo continuo de las mejoras instaladas, establecer los plazos de la expiración del modelo y los umbrales para redefinir la metodología estableciendo si el estudio realizado puede aplicarse en otras áreas de la entidad. (Israel, 2022).

- La generación del informe final en donde se involucran todas las áreas impactadas en el proyecto especialmente con los equipos de TI para garantizar

una implementación exitosa y asegurarse de que los resultados de la minería de datos sean utilizados de manera efectiva para beneficiar al CNMH. (Haya. 2021)

3. Resultados esperados

Se espera que, con los lineamientos otorgados en esta consultoría y la ejecución de cada uno de los pasos de la metodología de minería de datos seleccionada, se obtengan los siguientes resultados esperados para la mejora en la gestión de metas comprometidas en instancias de coordinación y articulación:

- Gestión de la información en el CNMH:
  Se pretende establecer los lineamientos estándar para el reporte, recolección y procesamientos de datos de la entidad. Así mismo, establecer y cumplir con el plan de capacitación del correcto ingreso y manejo de las bases de datos entregadas por la entidad a todas las personas involucradas en este rol.

- Inventarios de fuentes de información:
  Al crear el DataLake, se espera que todas las fuentes de información manejadas por el CNMH queden allí consignadas para evitar fuga de la información, siendo una guía importante para el departamento de TI.

- Uso de herramientas para la gestión y modelado de la información:
  Con el manejo de KNIME y/o Python se espera contar con una limpieza y transformación óptima de los datos para su posterior análisis.

- Implementación y creación de tableros de control
  Se mostrarán los resultados al CNMH en una herramienta de visualización, que en este caso sería Looker Studio, de una forma dinámica, de fácil entendimiento

e interacción para una compresión ágil y eficaz de las metas e indicadores en instancias de coordinación y articulación.

4. Conclusiones y recomendaciones

4.1 Conclusiones

La identificación de fuentes de datos y procesos en la gestión de metas e indicadores del Centro Nacional de Memoria Histórica es un paso fundamental para comprender la complejidad de los desafíos que enfrenta la entidad. La diversidad de fuentes y la estructuración en los procesos actuales destacan la necesidad de una estrategia cohesiva para mejorar la gestión de la información.

La propuesta de técnicas y herramientas de transformación de datos representa un avance significativo en el abordaje de los retos identificados. La aplicación de estas metodologías no solo mejorará la calidad de los datos, sino que también brindará las bases para la aplicación efectiva de técnicas de minería de datos en la consecución de los objetivos del CNMH.

La implementación de un plan de acción basado en la metodología CRISP-DM, es una estrategia efectiva para abordar los desafíos en la gestión de información. La aplicación de buenas prácticas en el procesamiento, organización y visualización de datos

contribuirá a la generación de información más significativa y útil para la toma de decisiones, mejorando así la visibilidad de la contribución del CNMH a la reparación a las víctimas y al derecho a la verdad en Colombia.

4.2 Recomendaciones

Optimizar los tiempos en la etapa de preparación de los datos, ya que es una de las que más demanda tiempo en la ejecución del proyecto. Por ello se recomienda la implementación de un sistema de recolección de información de fácil manejo y clasificación, que permita llevar a cabo el proceso de minería de datos de manera exitosa apuntando a contar con una información de calidad para la toma de decisiones en el CNMH.

Concretar reuniones con los involucrados del CMNH para establecer, entender y proyectar los diferentes gráficos y la forma en que desean visualizar los cambios de las metas indicadas. Así mismo, considerar si estas solicitudes van a afectar el modelo de minería de datos seleccionado.

Realizar auditorías a todos los procesos que intervienen en la ejecución de las metas, para confirmar que lo que se encuentra plasmado en los documentos entregados por el CNMH sean verídicos y contribuyan con las acciones propuestas por la entidad.

Bibliografía

Centro Nacional de Memoria Histórica. (2023). Consultoría especializada para el diseño del modelo de gestión del conocimiento y la innovación para el Centro Nacional de Memoria Histórica.

Centro Nacional de Memoria Histórica. (2020). Misión y Visión - Centro Nacional de Memoria Histórica. https://centrodememoriahistorica.gov.co/mision-y-vision/

Agencia Cooperación Alemana. (2022). PROPAZ II. Consolidación de la paz en Colombia https://www.giz.de/de/downloads/giz-2022-es-PROPAZ.pdf

Agencia Cooperación Alemana. (2023). Consultoría con el objeto: Elaborar los capítulos iniciales de marco conceptual, diagnóstico, antecedentes y justificación del documento CONPES de formulación de la política de verdad y memoria. Proyecto

PROPAZ II

Departamento Nacional de Planeación. (2022). Plan Nacional de Desarrollo 2022-2026. https://www.dnp.gov.co/plan-nacional-desarrollo/pnd-2022-2026

Google. (s. f.) Google Colab. https://research.google.com/colaboratory/intl/es/faq.html

Google Cloud. (s. f.). Looker Studio https://cloud.google.com/looker-studio?hl=es

Haya, P. (2021). La metodología CRISP-DM en ciencia de datos - IIC. Instituto de Ingeniería del Conocimiento. https://www.iic.uam.es/innovacion/metodologia-crisp-dm-ciencia-de-datos/

Hernández, R., Fernández, C., & Baptista, P. (2018). Metodología de la investigación (Vol. 4, pp. 310-386). México: McGraw-Hill Interamericana.

IBM. (2021). IBM Documentation. https://www.ibm.com/docs/es/spss-modeler/saas?topic=preparation-ready-modeling

IBM. (s. f.). Guia de CRISP- DM de IBM SPSS Modeler https://www.ibm.com/docs/es/SS3RA7_18.4.0/pdf/ModelerCRISPDM.pdf

Israel, R. (2022). CRISP-DM, Metodología de Datos. https://ipmoguide.com/crisp-dm-metodologia-de-datos/

KNIME Analytics Platform. (s. f.). KNIME. Obtenido de https://www.knime.com/knime-analytics-platform

Ley 1448 de 2011. (2016). Unidad para las Víctimas. 3 de Marzo de 2016 https://www.unidadvictimas.gov.co/es/documentos_bibliotec/ley-1448-de-2011-2/

Ludus, D. (2018). Administrar proyectos de data Crisp DM. Youtube. https://www.youtube.com/watch?v=QKq6yFZ76bw

Matplotlib.org. (s. f.). Matplotlib — visualization with python. https://matplotlib.org/

Moine, J. M., Gordillo, S., & Haedo, A. S. (2011). Análisis comparativo de metodologías para la gestión de proyectos de Minería de Datos. VIII Workshop Bases de Datos y Minería de Datos (WBDDM), 931–938.

Numpy.org. (s. f.). NumPy documentation — NumPy v1.26 manual. https://numpy.org/doc/stable/

Pydata.org. (s. f.). Pandas documentation — pandas 2.1.3 https://pandas.pydata.org/docs/

Pydata.org. (s. f.). Seaborn: Statistical data visualization — seaborn 0.13.0 https://seaborn.pydata.org/

Python.org. (s. f.). Welcome to. https://www.python.org/about/

Shafique, U., & Qaiser, H. (2014). A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA). International Journal of Innovation and Scientific Research, 12(1), 217– 222. http://www.ijisr.issr-journals.org/