



CLIAS

CENTRO DE INTELIGENCIA
ARTIFICIAL Y SALUD
PARA AMÉRICA LATINA
Y EL CARIBE

REPORTE FINAL

Detección automática de consultas para apoyar la atención de la salud sexual y reproductiva de personas con discapacidad.

INSTITUTO

CIECTI - Centro Interdisciplinario de Estudios en Ciencia, Tecnología e Innovación.

PERÍODO DEL INFORME

20/11/23 - 19/11/24





CLIAS

CENTRO DE INTELIGENCIA
ARTIFICIAL Y SALUD PARA
AMÉRICA LATINA Y EL CARIBE

INTEGRANTES DEL EQUIPO

Laura Alonso Alemany,
Marina Elichiry,
Sabrina López,
Cecilia Palermo,
Mariela Rajngewerc,
Milagro Teruel,
María Victoria Tiseyra,
Verónica Xhardez.

COLABORADORES

Melisa Tolaba,
Marcelo Pérez (+),
Carolina Buceta,
Karina Guershberg,
Eva Panaro,
Ximena Arriola,
Paula Schwartzman
Federico Céstars.



CONTENIDO

| | |
|--|----------|
| 01. INTRODUCCIÓN | 4 |
| 02. OBJETIVOS | 4 |
| 03. ACTIVIDADES REALIZADAS | 5 |
| 3.1 MESAS DE TRABAJO | 5 |
| 3.2 MANUAL DE ANOTACIÓN | 5 |
| 3.3 DESARROLLO DEL CONJUNTO DE DATOS ANOTADO | 5 |
| 3.4 ENTRENAMIENTO Y EVALUACIÓN DE MODELOS | 6 |
| 04. RESULTADOS | 6 |



01. INTRODUCCIÓN

En Argentina, la legislación vigente reconoce los derechos sexuales y reproductivos como derechos humanos básicos, garantizando la posibilidad de vivir la sexualidad de manera libre y sin discriminación. Sin embargo, las personas con discapacidad (PCD) enfrentan numerosos prejuicios y barreras que dificultan su acceso a una atención adecuada en materia de salud sexual y reproductiva (SSyR). Esta investigación se presenta como un análisis de factibilidad necesario para el desarrollo de aplicaciones integradas a historias clínicas electrónicas (HCE) que recomienden al equipo de salud – en las ocasiones donde resulte necesario– brindar recursos y apoyos adecuados y oportunos a las PCD destinatarias de la atención de SSyR. Además, la detección de consultas de este grupo poblacional permitiría proporcionar información a los niveles de mesogestión. Esto resulta valioso para un mejor direccionamiento de los recursos y una visibilización de la población de PCD para orientar las políticas en materia de SSyR, en un contexto de escasez de datos estadísticos sobre esta población.

02. OBJETIVOS

El objetivo principal de esta propuesta fue desarrollar una herramienta que permita la detección automática de consultas de Salud Sexual y Reproductiva (SSyR) de PCD a partir de modelos de inteligencia artificial (IA) aplicados a datos de HCE, que pueda ser un insumo para sistemas de alerta o de gestión de información.

Los objetivos específicos planteados fueron:

1. Constituir definiciones operativas de “consulta sobre salud sexual y/o reproductiva” y de “persona con discapacidad” en el contexto de HCE del primer nivel de atención.
2. Identificar elementos textuales (entidades) con información valiosa y otras características para la detección de las consultas y las personas objetivo tanto en el texto libre (evoluciones), como en la información estructurada en la Base de Datos.
3. Desarrollar un conjunto de datos en el que se identifique manualmente la información de interés para el entrenamiento y evaluación de un método automático de detección.
4. Obtener modelos de clasificación para identificar consultas de SSyR y de PCD.
5. Evaluar en profundidad los modelos de clasificación obtenidos, con especial atención a posibles problemas de discriminación, aplicando métricas de equidad específicas.



03. ACTIVIDADES REALIZADAS

3.1 MESAS DE TRABAJO

Se organizaron dos mesas de trabajo colaborativas con el objetivo de conocer e intercambiar experiencias con las comunidades involucradas en la investigación. La primera mesa se centró en el acceso a la salud, SSyR, y el registro de discapacidad en las historias clínicas, permitiendo identificar consensos y desacuerdos sobre estos temas. La segunda mesa se focalizó en discutir las definiciones conceptuales sobre discapacidad y SSyR vigentes y elaborar colaborativamente pautas para identificar registros de consultas de salud de PCD. Como resultado de estas discusiones se elaboró el documento: “Definiciones conceptuales y operativas de Personas con Discapacidad y de consultas de Salud Sexual y Reproductiva (V.2)”. Las mesas facilitaron el intercambio de vivencias y conocimiento experto y fomentaron la participación de PCD y de personas que trabajan en organismos e instituciones para PCD, logrando definiciones y criterios más inclusivos y precisos.

3.2 MANUAL DE ANOTACIÓN

Para guiar el proceso de anotación de datos, se elaboró un manual de anotación. Este documento proporciona directrices detalladas para los anotadores, asegurando la consistencia y la calidad de las anotaciones. Incluye definiciones claras de las categorías a anotar, ejemplos prácticos, y un conjunto de instrucciones paso a paso. Además, el manual contempla preguntas frecuentes y soluciones a posibles problemas que podrían surgir durante el desarrollo de esta tarea.

3.3 DESARROLLO DEL CONJUNTO DE DATOS ANOTADO

La base de datos original contenía registros de atención de pacientes del sistema público de atención primaria en La Rioja, Argentina, recopilados entre 2016 y 2022 y anonimizados para su uso secundario en investigación. Tras eliminar registros irrelevantes y aplicar un proceso de pretratamiento, se redujo significativamente la cantidad de datos. Luego, se identificaron casos relevantes relacionados con SSyR y PCD, los que fueron seleccionados para la anotación manual y se organizaron en categorías específicas.

El proceso de anotación manual contó con la colaboración de cinco anotadoras, siendo una de ellas PCD. Todas trabajaron con el apoyo del manual de anotación elaborado y participaron en rondas de consenso para garantizar una clasificación precisa. Las categorías de anotación abarcaban temas de SSyR y PCD. El conjunto de datos anotado



final consta de 3299 registros, con 2300 relevantes para SSyR y 2004 para PCD, de los cuales 1005 pertenecen a ambas categorías.

Las participantes resaltaron la facilidad de uso de la plataforma, la comunicación efectiva y el tiempo adecuado. Sin embargo, enfrentaron dificultades en la obtención de consensos y el impacto emocional de leer casos sensibles.

3.4 ENTRENAMIENTO Y EVALUACIÓN DE MODELOS

Se utilizaron varios enfoques de aprendizaje automático sobre el conjunto de datos anotado para el entrenamiento y evaluación de modelos. Se exploraron distintas representaciones del texto, como *tf-idf* y *word embeddings* especializados en el ámbito médico.

Los modelos utilizados incluyeron el Clasificador Bayesiano Ingenuo Multinomial, la Regresión Logística, la Máquina de Vector de Soporte y XGBoost. Cada uno fue optimizado mediante una búsqueda de hiperparámetros y evaluado con métricas como la tasa de aciertos, precisión y cobertura por clase.

Se seleccionó el modelo con mejor desempeño para la identificación automática de consultas relacionadas con salud sexual y reproductiva y discapacidad. Luego, se hizo un análisis de equidad (*fairness*) para estudiar si el modelo estaba generando errores sistemáticos en determinados grupos de la población.

04. RESULTADOS

Los modelos de aprendizaje automático desarrollados para la detección de consultas sobre SSyR y PCD lograron una tasa de acierto del 80% para cada categoría por separado y del 62% para la identificación simultánea de ambas. Estos resultados superan significativamente a los obtenidos con métodos basados en expresiones regulares y reglas lógicas. Sin embargo, se identificaron áreas de mejora en la detección de sesgos, y se han propuesto estrategias para abordarlas en el futuro. Durante el desarrollo del proyecto, el equipo tomó las precauciones necesarias para el manejo responsable de datos sensibles. Asimismo, participaron en eventos científicos relacionados con la temática de la investigación.

Es importante destacar, que se involucró activamente a la comunidad destinataria en la construcción de definiciones y en el proceso de anotación de datos. Los documentos elaborados, relacionados con las definiciones operativas y el manual de anotación, están disponibles.



CLIAS

CENTRO DE INTELIGENCIA
ARTIFICIAL Y SALUD
PARA AMÉRICA LATINA
Y EL CARIBE

