

Modelo de Segmentación de Riesgo en Morbilidad Materna Extrema mediante Técnicas de Aprendizaje No Supervisado en gestantes de Sucre, Colombia.

German Tercero Robles Monterroza

Asesor

Doctor Luis Ángel Anillo Arrieta

Universidad Nacional Abierta y a Distancia UNAD
Escuela de Ciencias Básicas Tecnologías e Ingeniería
Especialización en ciencia de datos y analítica

Octubre de 2025

Tabla de contenido

Introducción	6
Descripción del problema	8
Planteamiento del problema.....	8
Justificación	10
Objetivos	11
Objetivo General	11
Objetivos Específicos.....	11
Marco de referencia	12
Estado del arte.....	12
Marco Contextual.....	16
Marco teórico	18
Marco Conceptual.....	21
Marco normativo.....	22
Metodología	24
Método	24
Tipo de estudio.....	25
Recolección de datos.....	25
Resultados	26
Preparación y depuración de los datos.....	26
Modelos de agrupamiento aplicados.....	29
Comparativa de desempeño de los modelos	36
Interpretación de los clusters	37

Discusión.....	39
Conclusión del capítulo.....	42
Conclusiones.....	43
Recomendaciones	45
Referencias Bibliográficas	47

Lista de Tablas

Tabla 1 <i>Marco conceptual</i>	21
Tabla 2 <i>Variables analizadas, tipo de dato, descripción y valores posibles</i>	27
Tabla 3 <i>Métricas del modelo optimo encontrado con el algoritmo k-means</i>	30
Tabla 4 <i>Metricas del modelo optimo encontrado con el algoritmo DBSCAN</i>	32
Tabla 5 <i>Porcentaje de puntos asignados a cada cluster y ruido total (DBSCAN)</i>	33
Tabla 6 <i>Metricas del modelo optimo del algoritmo de clustering aglomerativo</i>	34
Tabla 7 <i>Promedios de variables por clúster</i>	35
Tabla 8 <i>Comparativa de las Métricas de evaluación de los modelos entrenados</i>	36
Tabla 9 <i>Comparativa de las Métricas de evaluación de los modelos entrenados</i>	38
Tabla 10 <i>Ventajas y limitaciones técnicas observadas en cada método</i>	41

Lista de Figuras

Figura 1	<i>Varianza explicada y acumulada por componente principal (scree plot)</i>	28
Figura 2	<i>Scree plot para la selección del número óptimo de clústeres (Inercia vs k)....</i>	29
Figura 3	<i>Distribución bidimensional de los clústeres obtenidos con K-Means (PCA) ..</i>	31
Figura 4	<i>Dendograma del agrupamiento jerárquico (ward linkage)</i>	34
Figura 5	<i>Comparativa gráfica de las métricas de desempeño entre modelos</i>	37
Figura 6	<i>Boxplots comparativos de edad, y número de controles prenatales por cluster</i>	39

Introducción

Este proyecto tiene como objetivo principal desarrollar un modelo para identificar patrones y segmentar grupos de riesgo de morbilidad materna extrema a partir del análisis de información clínica, socioeconómica y de acceso a servicios de salud. A diferencia de los modelos predictivos tradicionales, este estudio se fundamenta en el aprendizaje no supervisado, lo que permite descubrir estructuras latentes y relaciones ocultas en los datos sin requerir etiquetas predefinidas. Se pretende que sirva para anticipar la morbilidad materna extrema (MME) en el departamento de Sucre, Colombia. Esta condición representa un problema de salud pública de gran magnitud, especialmente entre las mujeres en situación de vulnerabilidad que habitan zonas rurales e indígenas, donde el acceso oportuno y de calidad a los servicios de salud es limitado.

A partir de la integración de datos clínicos, socioeconómicos y de acceso a servicios de salud, se buscará identificar los principales factores de riesgo que predisponen a complicaciones durante el embarazo, el parto y el puerperio. La propuesta metodológica contempla el uso de técnicas de reducción de dimensionalidad y la aplicación de distintos algoritmos de aprendizaje no supervisado, tales como K-Means, DBSCAN y Clustering aglomerativo, evaluados mediante métricas como el Coeficiente de silueta, Índice de Davies–Boulding y Calinski-Harabaz.

Este trabajo tiene un enfoque aplicado y orientado al impacto. Se espera que los resultados contribuyan significativamente a la detección temprana de gestantes en riesgo, fortaleciendo así las capacidades del sistema de salud en Sucre y favoreciendo la toma de decisiones clínicas y de política pública. En última instancia, se busca aportar al cumplimiento de

los Objetivos de Desarrollo Sostenible mediante la reducción de complicaciones graves y la mejora de la salud materna.

La presente memoria se estructura de la siguiente manera: Se presenta el contexto del problema y una revisión del estado del arte sobre el uso de técnicas de minería de datos y aprendizaje no supervisado en salud materna. Luego se expone los objetivos y la metodología adoptada. Posteriormente se describe el proceso de análisis, selección de variables y preprocesamiento de datos. Luego aborda la implementación de los modelos de segmentación y la interpretación de los resultados, comparación de los modelos con la correspondiente discusión y finalmente, presenta las conclusiones y recomendaciones.

Descripción del problema

Planteamiento del problema

El presente proyecto tiene como propósito desarrollar un modelo de segmentación basado en técnicas de aprendizaje automático que permita brindar herramientas para prevenir y anticipar la ocurrencia de morbilidad materna extrema (MME) en las gestantes del departamento de Sucre, Colombia. La finalidad de este sistema es contribuir a la identificación temprana de mujeres con perfil de riesgo, fortaleciendo la capacidad de respuesta del sistema de salud y apoyando la toma de decisiones clínicas y de política pública.

La morbilidad materna extrema (MME) se define como una complicación grave que ocurre durante el embarazo, el parto o dentro de los 42 días posteriores a la terminación del embarazo, que pone en riesgo la vida de la mujer, pero de la cual sobrevive gracias a una intervención médica oportuna (Instituto Nacional de Salud [INS], 2024). Este evento es un indicador crucial de la calidad de la atención obstétrica y refleja las desigualdades existentes en el sistema de salud (Campaña Bastidas et al., 2024)

En el departamento de Sucre, la situación es particularmente preocupante. Según el Instituto Nacional de Salud de Colombia, en 2021 se registró una razón de morbilidad materna extrema (MME) superior a 47,2 casos por cada 1.000 nacidos vivos, mientras que a nivel nacional la razón fue de 48,4 casos por cada 1.000 nacidos vivos. Además, en 2023 se reportaron 33 casos de MME en Sucre, lo que representa el 2,25% de los eventos notificados en la región. Estas cifras reflejan desigualdades significativas en el acceso y la calidad de la atención en salud, especialmente en zonas rurales e indígenas del departamento.

Es crucial abordar este problema, ya que cada caso de morbilidad materna extrema tiene un impacto devastador tanto a nivel individual como social. La pérdida o discapacidad de la

madre afecta el desarrollo de los niños, aumenta el riesgo de pobreza y perpetúa desigualdades en salud y acceso a servicios. Además, este problema refleja el grado de desarrollo del sistema de salud y tiene implicaciones directas para el cumplimiento de los Objetivos de Desarrollo Sostenible, en particular, la meta de reducir la razón de mortalidad materna.

Las causas de la morbilidad materna extrema son múltiples y abarcan factores médicos y sociales. Entre ellas se incluyen la falta de acceso a atención prenatal de calidad, desigualdades en el sistema de salud, complicaciones obstétricas no tratadas adecuadamente (como hemorragias o infecciones), condiciones preexistentes de la madre y deficiencias en la identificación temprana de riesgos. La situación se agrava en áreas con menos infraestructura y personal capacitado, así como entre poblaciones vulnerables, como mujeres indígenas o migrantes.

Si no se abordan de manera efectiva la morbilidad y mortalidad materna, se perpetuarán los ciclos de pobreza y desigualdad. El impacto negativo en los hogares es significativo, con niños huérfanos que enfrentan peores condiciones de vida y desarrollo. Además, la falta de progreso en este indicador de salud obstaculizará los avances hacia los compromisos internacionales de reducción de la mortalidad materna para 2030, como parte de los Objetivos de Desarrollo Sostenible.

El problema afecta desproporcionadamente a mujeres de comunidades rurales, indígenas y en situación de pobreza, donde el acceso a servicios de salud es limitado. La alta morbilidad materna extrema también incrementa los costos del sistema de salud debido a hospitalizaciones prolongadas y cuidados intensivos. A nivel social, se traduce en pérdidas económicas y emocionales significativas para las familias y comunidades enteras, perpetuando la vulnerabilidad económica y limitando las oportunidades de desarrollo para las futuras generaciones.

Justificación

En Colombia, la morbilidad materna extrema (MME) continúa siendo un desafío significativo para la salud pública. Según el Instituto Nacional de Salud, en 2020 se registraron 24.451 casos de MME, lo que representa una razón de 39,5 casos por cada 1.000 nacidos vivos, evidenciando un incremento del 5,8% en comparación con el año anterior. (Instituto nacional de salud, 2024).

A pesar de los esfuerzos implementados, existe una notable escasez de estudios que aborden la MME en el contexto colombiano, especialmente aquellos que empleen enfoques innovadores como el Machine Learning para la predicción de riesgos. Esta carencia de investigaciones limita la capacidad para desarrollar herramientas efectivas que permitan la identificación temprana de mujeres en riesgo y la implementación de intervenciones oportunas.

El desarrollo de un modelo estratificación de riesgos basado en Machine Learning que incorpore factores clínicos, socioeconómicos y de acceso a servicios de salud se alinea directamente con las políticas de salud pública del país. El Ministerio de Salud y Protección Social, en el marco de la Política de Atención Integral en Salud (PAIS) y la aplicación del Modelo de Atención Integral en Salud (MIAS), busca garantizar una atención integral a las gestantes, sus familias y comunidades, mediante intervenciones de valoración y manejo del riesgo. (Ministerio de salud y protección social, 2024)

Este proyecto contribuiría a fortalecer las estrategias nacionales dirigidas a reducir la morbilidad y mortalidad materna, proporcionando una herramienta basada en evidencia que facilite la toma de decisiones clínicas y de salud pública, mejorando así los resultados en salud materna y avanzando hacia el cumplimiento de los objetivos establecidos en las políticas nacionales.

Objetivos

Objetivo General

Desarrollar un modelo de estratificación de riesgo utilizando técnicas de Aprendizaje No Supervisado para identificar perfiles latentes de Morbilidad Materna entre las gestantes del departamento de Sucre.

Objetivos Específicos

- Realizar una revisión del estado del arte de estudios previos en predicción de morbilidad materna extrema, identificando metodologías y factores de riesgo clave reportados.
- Recolectar datos de información clínica (hipertensión, diabetes, IMC), socioeconómica (nivel educativo, ocupación) y acceso a servicios de salud (distancia al centro de salud, visitas prenatales). De embarazadas incluyendo complicaciones en el embarazo, parto o puerperio luego de 42 días posterior al parto.
- Implementar técnicas reducción de dimensionalidad para encontrar las variables que inciden en la presentación o no de Morbilidad Materna extrema.
- Entrenar diversos algoritmos de Machine Learning No supervisado para identificar patrones en los datos y agruparlos por niveles de riesgo de Morbilidad Materna Extrema.
- Evaluar los modelos en términos de cohesión de los grupos encontrados en los datos. Mediante métricas como el Coeficiente de silueta, Índice de Davies–Bouldin y Calinski-Harabaz.

Marco de referencia

Estado del arte

El presente proyecto se enmarca en la creciente intersección entre la Salud Materna y la Inteligencia Artificial (IA), específicamente el Aprendizaje Automático (Machine Learning, ML), con el objetivo de predecir la Morbilidad Materna Extrema (MME). La literatura reciente demuestra una tendencia global hacia la aplicación de modelos predictivos basados en datos para la detección temprana de complicaciones obstétricas graves.

El análisis de los referentes seleccionados revela tres pilares fundamentales en la investigación actual: la aplicación metodológica del ML en obstetricia, la predicción específica de complicaciones de alto riesgo (como preeclampsia y sepsis), y la pertinencia del enfoque en contextos de países de ingresos medios.

- *Aplicación Metodológica del Machine Learning en Obstetricia*

Una revisión sistemática clave de Islam et al. (2022) establece un marco conceptual para el uso de ML en la predicción de resultados del embarazo. Este estudio destaca que el ML se utiliza para determinar el modo óptimo de parto y detectar complicaciones, analizando la evolución de los objetivos de investigación, los algoritmos y las características de los datos a lo largo del tiempo. Un aporte crucial para el proyecto es la identificación de futuras oportunidades de investigación, incluyendo el uso de algoritmos no supervisados (que se alinea con el Objetivo Específico 4 del proyecto) y el desarrollo de sistemas de apoyo clínico (Islam et al., 2022).

En cuanto a las técnicas específicas, Lakshmi et al. (2016) demuestran la efectividad del algoritmo de clasificación de árboles de decisión C4.5 para predicciones de riesgo durante el embarazo. Su principal conclusión subraya la importancia crítica de la estandarización de los

parámetros de recolección de datos para mejorar la precisión predictiva, un punto directamente relevante para el trabajo con registros de salud colombianos que pueden ser incompletos o no estandarizados.

Complementariamente, el estudio de Raza et al. (2022) propone una arquitectura avanzada de red neuronal profunda (DT-BiLTCN) utilizando ingeniería de características basada en *ensemble learning* para predecir riesgos en la salud materna. Este trabajo no solo muestra la capacidad predictiva de los modelos de IA (alcanzando una precisión del 98%) sino que también resalta variables indicativas de riesgo, como la presión arterial y la frecuencia cardíaca (Raza et al., 2022).

- *Predicción Específica de Complicaciones de Alto Riesgo*

La literatura se enfoca significativamente en la predicción temprana de dos de las principales causas de MME: la preeclampsia y la sepsis materna. La preeclampsia es una de las principales causas de morbilidad materno-fetal, y los referentes abordan su predicción desde diferentes ángulos; Maric et al. (2020) proponen un enfoque no invasivo y accesible para detectar la PE, utilizando mediciones de presión arterial obtenidas mediante una pulsera inteligente. Esto subraya la viabilidad de implementar soluciones de monitoreo continuo en entornos remotos o de difícil acceso, como las zonas rurales de Sucre; Olano et al. (2023) desarrollan un árbol de clasificación predictivo de PE en embarazos de alto riesgo basado en variables hemodinámicas no invasivas. El modelo demostró una alta precisión (93,75%), validando el uso de parámetros fisiológicos sencillos como predictores tempranos. Torres-Torres et al. (2023) enfocan su trabajo en países de ingresos medios, como México, demostrando que un modelo ML puede lograr una alta precisión en la predicción de PE temprana utilizando características maternas del primer trimestre y valores derivados localmente (MoM). Este es un

referente clave, ya que la aplicabilidad en México sugiere una potencial adaptación exitosa a un contexto socioeconómico y sanitario similar al colombiano. Gonzales Medina y Alegría Guerrero (2014) ofrecen una revisión integral sobre la predicción de PE, destacando la importancia de la flujometría Doppler, biomarcadores bioquímicos y factores clínicos asociados. Este referente es fundamental para guiar la selección de variables clínicas en la fase de recolección de datos.

- *Sepsis Materna y Trastornos Hipertensivos*

La sepsis materna es otra causa importante de MME. Kopanitsa et al. (2021) desarrollan un modelo de ML que predice el riesgo de sepsis en el embarazo utilizando signos vitales y antecedentes de la mujer, reduciendo la necesidad de pruebas de laboratorio costosas. La simplicidad de las variables de entrada propuestas por este modelo aumenta su viabilidad en contextos con recursos limitados.

En cuanto a los trastornos hipertensivos, Nielsen et al. (2023) proponen una plataforma con IA y sensores biométricos para el monitoreo continuo y ambulatorio de la presión arterial. Este enfoque tecnológico es esencial para el manejo proactivo y preventivo de una de las principales causas de MME, como lo mencionan Campaña Bastidas et al. (2024) para el caso colombiano.

- *Contexto nacional y la pertinencia del estudio para Colombia*

El proyecto encuentra un precedente directo en la investigación de Campaña Bastidas et al. (2024), que describe el avance en el desarrollo de un sistema tecnológico con ML para la prevención de MME en la Fundación Hospital San Pedro de Pasto, Colombia. Este estudio valida la necesidad y la aplicabilidad de las herramientas de IA en el sistema de salud colombiano, enfocándose en la caracterización de la población gestante y la identificación de patrones de riesgo relacionados con trastornos hipertensivos y complicaciones hemorrágicas.

La revisión sistemática de Islam et al. (2022) ya había anticipado esta adaptabilidad, señalando que el marco conceptual de ML puede integrarse con variables locales específicas, como características socioeconómicas, geográficas y de acceso a servicios de salud, lo cual constituye el corazón de la pregunta problema del proyecto.

En síntesis, el estado del arte valida la viabilidad técnica (ML, Redes Neuronales, C4.5) y la relevancia clínica (predicción de PE y sepsis) del proyecto, y confirma su pertinencia contextual al existir esfuerzos similares en el país que buscan integrar la información clínica, socioeconómica y de acceso a la salud para la predicción de MME en Colombia.

Marco Contextual

El proyecto se ubica geográficamente en el Departamento de Sucre, Colombia, una región que presenta desafíos significativos en salud materna que hacen imperativo el uso de herramientas de analítica de datos innovadora innovadoras.

- *Morbilidad Materna Extrema en Sucre*

Cómo se presentó en el apartado de la justificación del proyecto, el departamento de Sucre enfrenta una situación preocupante respecto a la Morbilidad Materna Extrema (MME), con 33 casos reportados en 2023, lo que representa el 2,25% de los eventos notificados en la región. Aunque la razón de MME de 47,2 casos por 1.000 nacidos vivos en 2021 fue ligeramente inferior al promedio nacional (48,4/1.000), estas cifras ocultan desigualdades intrarregionales críticas.

- *Desafíos Socioeconómicos y de Acceso a la Salud*

Sucre, como muchos departamentos de la Costa Caribe colombiana, se caracteriza por la dispersión geográfica de su población, la presencia de comunidades rurales e indígenas, y notorias limitaciones en la infraestructura y el acceso oportuno a servicios de salud de alta complejidad. La justificación y la introducción del proyecto resaltan que la MME impacta desproporcionadamente a mujeres en situación de vulnerabilidad, lo que exige que el modelo predictivo incorpore, además de las variables clínicas, factores socioeconómicos (nivel educativo, ocupación) y de acceso a servicios de salud (distancia al centro de salud, visitas prenatales).

- *Necesidad de un Modelo Predictivo Basado en IA*

El contexto de Sucre es ideal para la aplicación de un modelo de ML. Ante la limitación de recursos y la dificultad de acceso a centros especializados, una herramienta predictiva (similar

a la pulsera inteligente de Marić et al. (2020) o la plataforma de monitoreo de Nielsen et al. (2023)) podría:

- Priorizar la atención: Identificar a las gestantes de más alto riesgo para garantizar su traslado y monitoreo más riguroso.
- Optimizar recursos: Enfocar los limitados recursos de atención especializada hacia donde son más necesarios, apoyando el sistema de salud colombiano, el cual busca un Modelo de Atención Integral en Salud (MIAS) (Ministerio de salud y protección social, 2024).
- Facilitar la toma de decisiones: Brindar a los profesionales de la salud en el primer nivel de atención (el más accesible en zonas rurales) una herramienta para la intervención temprana, como lo propone el estudio de Campaña Bastidas et al. (2024) para Pasto.

El proyecto, al enfocarse en Sucre, se convierte en un estudio de caso crucial sobre cómo la tecnología predictiva puede mitigar las inequidades en salud en contextos de recursos limitados en Colombia.

Marco teórico

El marco teórico del proyecto se sustenta en tres áreas disciplinares clave: la Epidemiología Obstétrica, la Teoría de Sistemas de Clasificación de Riesgo y los Principios del Aprendizaje Automático.

- *Epidemiología Obstétrica: La Morbilidad Materna Extrema (MME)*

La Morbilidad Materna Extrema (MME), o *near-miss* materno, se define operacionalmente como una complicación severa que ocurre durante el embarazo, el parto o dentro de los 42 días siguientes a la terminación, que pone en riesgo la vida de la mujer, pero que es mitigada por una intervención médica oportuna y sobrevive (Instituto Nacional de Salud [INS], 2024).

- *Importancia:* La MME es un indicador de la calidad de la atención obstétrica más sensible que la mortalidad materna, ya que es más frecuente y permite identificar fallas en el sistema de salud que, de no corregirse, resultarían en muertes maternas.
- *Causas Principales:* El proyecto se enfoca en la predicción de las principales causas de MME en Colombia, que según Campaña Bastidas et al. (2024), son los trastornos hipertensivos relacionados con el embarazo (preeclampsia/eclampsia) y las complicaciones hemorrágicas.
- *Factores de Riesgo:* La teoría establece que la MME es el resultado de la interacción de factores biológicos (clínicos), factores sociales (socioeconómicos, étnicos) y factores relacionados con el sistema de salud (acceso, calidad de la atención). El modelo propuesto integra estos tres dominios como variables independientes.
- Teoría de Sistemas de Clasificación de Riesgo

El proyecto se basa en la necesidad teórica de mover el paradigma de la atención reactiva a la atención proactiva y preventiva. La teoría de la clasificación de riesgo en salud busca estratificar a la población en función de su probabilidad de experimentar un evento adverso.

- *Enfoque Predictivo*: El objetivo no es solo describir qué mujeres experimentan MME, sino predecir quién la experimentará, permitiendo una intervención temprana y la asignación de recursos.
- *Multifactorialidad*: El riesgo materno es inherentemente multifactorial. La incorporación de variables clínicas (como presión arterial o historia obstétrica), socioeconómicas (educación, pobreza) y de acceso a la salud es una base teórica sólida, respaldada por Islam et al. (2022), que señalan la necesidad de adaptar estos modelos a las variables locales.
- *Principios del Aprendizaje Automático (Machine Learning)*

El proyecto utiliza el ML como la herramienta analítica para construir el modelo predictivo.

- *Machine Learning (ML)*: Es una rama de la Inteligencia Artificial que dota a las computadoras de la capacidad de aprender a partir de datos, sin ser explícitamente programadas. El proceso de aprendizaje consiste en encontrar patrones y relaciones en un conjunto de datos para hacer predicciones o tomar decisiones sobre nuevos datos.
- *Aprendizaje No Supervisado (Clustering)*: El proyecto inicialmente utilizará algoritmos No Supervisados (Objetivo Específico 4). Teóricamente, el clustering (agrupamiento) se utiliza para descubrir patrones o estructuras ocultas en los datos de embarazos sin una etiqueta de resultado predefinida. Esto es útil para:

- ✓ *Segmentación de Riesgo:* Agrupar a las gestantes en perfiles de riesgo latentes (por ejemplo, un grupo de alto riesgo con baja educación y alta distancia al centro de salud, distinto a un grupo con solo hipertensión), lo cual se evaluará con métricas como el Coeficiente de Silueta.
- ✓ *Reducción de Dimensionalidad:* Técnicas como PCA (Análisis de Componentes Principales), mencionadas en el Objetivo Específico 2, buscan proyectar los datos en un espacio de menor dimensión manteniendo la máxima varianza, lo que simplifica y mejora la eficiencia de los modelos, un principio común en el *feature engineering* como el utilizado por Raza et al. (2022).
- *Validación del Modelo:* La teoría de la evaluación de modelos de ML exige el uso de métricas rigurosas. Las métricas de *clustering* propuestas (Coeficiente de Silueta, Davies-Bouldin, Calinski-Harabaz) miden la cohesión (qué tan unidos están los puntos dentro de un clúster) y la separación (qué tan lejos están los clústeres entre sí), asegurando que los grupos de riesgo identificados sean estadísticamente significativos.

Marco Conceptual

El marco conceptual define los términos clave del proyecto para asegurar una comprensión unívoca de los objetivos.

Tabla 1

Marco conceptual

Concepto	Definición para el Proyecto	Referencias de soporte
Morbilidad Materna Extrema (MME)	Complicación severa durante el embarazo, parto o puerperio (hasta 42 días postparto) que pone en riesgo la vida de la mujer, pero de la cual sobrevive gracias a una intervención médica oportuna. El resultado del modelo (variable dependiente).	INS (2024); Campaña Bastidas et al. (2024)
Aprendizaje Automático (Machine Learning, ML)	Rama de la IA que utiliza algoritmos para aprender patrones de los datos (clínicos, socioeconómicos, de acceso) para hacer predicciones sobre la ocurrencia o no de MME.	Islam et al. (2022); Raza et al. (2022)
Modelo Predictivo	Un algoritmo o función matemática entrenada con datos históricos para estimar la probabilidad de que una gestante desarrolle MME (clasificación binaria: MME sí/no).	Olano et al. (2023); Torres-Torres et al. (2023)
Preeclampsia (PE)	Trastorno hipertensivo del embarazo (una de las principales causas de MME). Se utiliza como un proxy clínico de alto riesgo para el entrenamiento y análisis del modelo.	Maric et al. (2020); Olano et al. (2023); Gonzales Medina y Alegría Guerrero (2014)
Algoritmos No Supervisados	Técnicas de ML (como clustering) que se entrenan sin etiquetas de salida predefinidas, buscando descubrir estructuras o grupos inherentes en el conjunto de datos de gestantes (ej. perfiles de riesgo).	Islam et al. (2022); Ob
Reducción de Dimensionalidad	Conjunto de técnicas (ej. PCA) para reducir el número de variables analizadas, eliminando la redundancia y manteniendo la información relevante para encontrar las que inciden en la MME (Objetivo Específico 2).	Islam et al. (2022) Lakshmi et al. (2016)
Acceso a Servicios de Salud	Variable socio-geográfica que mide la disponibilidad y facilidad de uso de los servicios de atención prenatal y obstétrica, como la distancia al centro de salud y el número de controles prenatales.	Islam et al. (2022) ; Campaña Bastidas et al. (2024)

Marco normativo

El proyecto se desarrolla dentro del sistema de salud colombiano y debe alinearse con la normatividad vigente, especialmente aquella relacionada con la vigilancia en salud pública, la protección de datos personales y la política de salud materna.

- **Vigilancia en Salud Pública y Morbilidad Materna Extrema**

Protocolo de Vigilancia en Salud Pública: La MME es un evento de interés en salud pública de notificación obligatoria en Colombia. El proyecto debe adherirse al Protocolo de Vigilancia de Morbilidad Materna Extrema establecido por el Instituto Nacional de Salud (INS), que define el evento, los criterios de notificación y las principales causas (trastornos hipertensivos y hemorrágicos). Esto asegura la coherencia en la definición de la variable dependiente (MME) utilizada en el entrenamiento del modelo.

Relevancia: Los datos recolectados (Objetivo Específico 1) deben provenir de o ser comparables con los registros del SIVIGILA (Sistema de Vigilancia Epidemiológica), garantizando la validez de la información epidemiológica.

- **Política de Atención Integral en Salud**

Ley 1751 de 2015 (Ley Estatutaria de Salud): Establece la salud como un derecho fundamental, lo que exige al Estado garantizar la accesibilidad y la calidad de la atención, especialmente para poblaciones vulnerables (como las gestantes en Sucre).

Política de Atención Integral en Salud (PAIS) y Modelo de Atención Integral en Salud (MIAS): El Ministerio de Salud y Protección Social promueve un modelo de atención que se centra en la gestión y manejo del riesgo (Ministerio de salud y protección social, 2024).

Relevancia: El modelo predictivo de ML contribuye directamente a la implementación del MIAS, al proporcionar una herramienta de valoración del riesgo y apoyo a la toma de decisiones clínicas y de salud pública.

- **Protección de Datos Personales y Ética en Investigación**

Ley 1581 de 2012 (Protección de Datos Personales): Cualquier recolección, almacenamiento y procesamiento de datos personales de las gestantes debe cumplir con los principios de legalidad, finalidad, libertad, veracidad, transparencia, acceso y circulación restringida, y seguridad.

Resolución 8430 de 1993 (Normas Científicas, Técnicas y Administrativas para la Investigación en Salud): El proyecto debe ser clasificado éticamente y, al trabajar con datos de pacientes, requerirá la aprobación de un Comité de Ética en Investigación de las instituciones de salud involucradas en Sucre.

Relevancia: La metodología debe garantizar la anonimización de los datos (como se menciona en la recolección de datos), asegurando que las variables de identificación no sean rastreables, cumpliendo así con los requisitos éticos y legales. La confidencialidad es crítica, dado que se utilizan datos clínicos, socioeconómicos y de geolocalización sensible.

Metodología

Este estudio utilizará una metodología cuantitativa, aplicada y predictiva, sustentada en técnicas de aprendizaje automático. Se parte del análisis de una base de datos retrospectiva de mujeres embarazadas del departamento de Sucre (Colombia), que incluye tanto casos normales como eventos de morbilidad materna extrema (MME) y mortalidad materna.

La metodología comprende:

- *Preparación y limpieza de los datos:* tratamiento de valores faltantes, detección y manejo de valores atípicos, y normalización de variables.
- *Análisis exploratorio y selección de características:* para identificar patrones relevantes.
- *Modelado de estratificación de riesgo* mediante algoritmos de aprendizaje automático no supervisado.
- *Evaluación de desempeño* de los modelos, con métricas como Coeficiente de silueta, Índice de Davies–Bouldin y Calinski-Harabaz.

Se buscará identificar los factores clínicos, socioeconómicos y de acceso a servicios de salud que más inciden en la aparición de MME.

Método

El método adoptado es cuantitativo, retrospectivo y no experimental, basado en el análisis de datos secundarios provenientes de registros clínicos y bases de datos institucionales. Se utilizará un enfoque de aprendizaje automático no supervisado (como K-Means, DBSCAN y Clustering Aglomerativo para explorar patrones latentes o segmentación de perfiles de riesgo).

La implementación técnica se realizará en Python utilizando bibliotecas como Pandas, Scikit-learn y Matplotlib/Seaborn. Se dividirán los datos en conjuntos de entrenamiento y prueba (train/test split o validación cruzada) para asegurar la validez del modelo.

Tipo de estudio

Se trata de un estudio analítico, observacional, retrospectivo y de tipo predictivo, en el que se utilizarán registros de embarazos pasados para generar un modelo que permita la segmentación de perfiles de riesgo de presentar morbilidad materna extrema a partir de variables clínicas, socioeconómicas y de acceso a la atención médica.

Recolección de datos

Los datos serán recolectados a partir de bases de datos secundarias, principalmente del Sistema de Vigilancia Epidemiológica de Colombia (SIVIGILA) y registros clínicos de instituciones de salud del departamento de Sucre.

Se priorizará la obtención de información de mujeres que hayan presentado complicaciones graves o mortalidad materna.

Las variables para recolectar incluyen:

- Datos clínicos: hipertensión, diabetes, IMC, edad materna, historia obstétrica.
- Datos socioeconómicos: escolaridad, ocupación, ingresos.
- Acceso a salud: distancia al centro médico, número de controles prenatales, atención especializada.
- Los datos serán anonimizados para preservar la confidencialidad y cumplir con las normas éticas y legales.

Resultados

Preparación y depuración de los datos

El conjunto de datos utilizado estuvo compuesto por registros clínicos anonimizados de gestantes atendidas en el departamento de Sucre (Colombia) obtenidos de fuentes secundarias como el Sistema de Vigilancia Epidemiológica (SIVIGILA). Cada registro corresponde a la notificación de un caso de Morbilidad materna extrema desde el año 2021 al año 2025.

Se implementó un riguroso proceso de depuración y selección de características (Feature Selection) para asegurar que el modelo se basara en patrones clínicos y no en ruido administrativo:

- *Eliminación de Variables Administrativas:* Columnas con alta cardinalidad y bajo valor predictivo para la severidad clínica (código de UPGD, municipio, departamento, aseguradora y ocupación) fueron eliminadas para aislar el patrón de riesgo intrínseco.
- *Imputación y Normalización:* Se realizó la imputación de valores faltantes y la estandarización z-score de las variables numéricas.
- *Reducción de Dimensionalidad:* se llevó a cabo aplicando el Análisis de Componentes Principales (PCA) sobre el conjunto de variables limpias y codificadas mediante *One-Hot Encoding*. Inicialmente, el espacio de características contaba con 72 columnas tras los procesos de limpieza y transformación de las variables. El objetivo de aplicar PCA fue mitigar el riesgo de sobreajuste (overfitting) y optimizar la eficiencia computacional del modelo posterior.

Para determinar el número óptimo de componentes a retener, se utilizó el criterio de varianza explicada acumulada, seleccionando el umbral que retuviera al menos el 90% de la varianza total de los datos originales.

El análisis de la curva de codo (Figura 1) indicó que el punto de inflexión necesario para alcanzar este umbral de varianza se situaba en 14 componentes principales (PC). Esta selección transformó el espacio de 72 dimensiones a un espacio latente de 14 dimensiones, logrando una significativa simplificación del modelo sin sacrificar una cantidad crítica de información contenida en los datos.

Tabla 2

Variables analizadas, tipo de dato, descripción y valores posibles

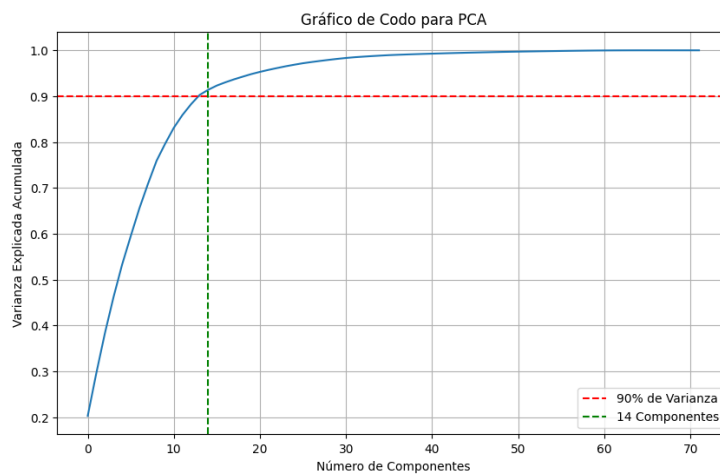
Variable	Tipo de Dato	Descripción	Valores Aceptados
edad_	Numérica discreta	Edad de la paciente en años.	Números enteros
num_gestac	Numérica discreta	Número total de gestaciones previas.	Números enteros (≥ 0)
num_parvag	Numérica discreta	Número de partos vaginales previos.	Números enteros (≥ 0)
num_cesare	Numérica discreta	Número de cesáreas previas.	Números enteros (≥ 0)
num_aborto	Numérica discreta	Número de abortos previos.	Números enteros (≥ 0)
num_molas	Numérica discreta	Número de gestaciones molares previas.	Números enteros (≥ 0)
num_ectopi	Numérica discreta	Número de embarazos ectópicos previos.	Números enteros (≥ 0)
num_muerto	Numérica discreta	Número de mortinatos o fetos muertos.	Números enteros (≥ 0)
num_vivos	Numérica discreta	Número de hijos vivos.	Números enteros (≥ 0)
no_con_pre	Catégorica Binaria	Indicador de si no tuvo control prenatal.	(0 = Sí tuvo, 1 = No tuvo)
sem_c_pren	Numérica discreta	Número de semanas de gestación al inicio del control prenatal.	Números enteros (≥ 0)
term_gesta	Catégorica nominal	Causa o razón de terminación de la gestación.	Texto catégorico (e.g., 'Parto', 'Aborto', 'Muerte Fetal')
moc_rel_tg	object	Motivo o condición relacionada con la terminación de la gestación.	Texto catégorico descriptivo.
falla_card	Catégorica Binaria	Indicador de falla o disfunción cardíaca.	(0 = No, 1 = Sí)
falla_rena	Catégorica Binaria	Indicador de falla o disfunción renal.	(0 = No, 1 = Sí)
falla_hepa	Catégorica Binaria	Indicador de falla o disfunción hepática.	(0 = No, 1 = Sí)

falla_cere	Catagórica Binaria	Indicador de falla o disfunción cerebral/neurológica.	(0 = No, 1 = Sí)
falla_resp	Catagórica Binaria	Indicador de falla o disfunción respiratoria.	(0 = No, 1 = Sí)
falla_coag	Catagórica Binaria	Indicador de falla o coagulopatía (problemas de coagulación).	(0 = No, 1 = Sí)
eclampsia	Catagórica Binaria	Indicador de Eclampsia.	(0 = No, 1 = Sí)
preclamsi	Catagórica Binaria	Indicador de Preeclampsia.	(0 = No, 1 = Sí)
choq_septi	Catagórica Binaria	Indicador de Choque séptico.	(0 = No, 1 = Sí)
hemorragia_obstétrica_severa	Catagórica Binaria	Indicador de Hemorragia Obstétrica Severa.	(0 = No, 1 = Sí)
rupt_uteri	Catagórica Binaria	Indicador de Ruptura Uterina.	(0 = No, 1 = Sí)
cir_adicio	Catagórica Binaria	Indicador de cirugías adicionales (histerectomía, etc.).	(0 = No, 1 = Sí)
ttl_criter	Numérica discreta	Total, de criterios o fallas cumplidas.	Números enteros (≥ 0)
dias_hospi	Numérica discreta	Días de hospitalización o estancia.	Números enteros (≥ 0)
caus_princ	Catagórica Nominal	Causa principal de morbilidad/ingreso (código o descripción).	Texto catagórico.
caus_agrup	Catagórica Nominal	Agrupación de la causa principal (catagoría mayor).	Texto catagórico.

Nota. La tabla presenta el Diccionario de Datos, que detalla las 29 variables usadas. Estas se dividen en 12 variables numéricas discretas (incluyendo edad, paridad y severidad de MME) y 17 variables catagóricas, la mayoría de las cuales son binarias (0/1) e indican la presencia de fallas orgánicas y complicaciones clínicas (p. ej., falla_card o eclampsia).

Figura 1

Varianza explicada y acumulada por componente principal (scree plot)



Fuente: Elaboración propia

Modelos de agrupamiento aplicados

Con el objetivo de identificar patrones latentes y posibles perfiles de riesgo de morbilidad materna extrema (MME), se aplicaron tres técnicas de aprendizaje no supervisado: K-Means, DBSCAN y Agrupamiento jerárquico aglomerativo.

Cada modelo fue evaluado mediante métricas no supervisadas de calidad de agrupamiento:

Coeficiente de Silueta, Índice de Davies–Bouldin e Índice de Calinski–Harabasz.

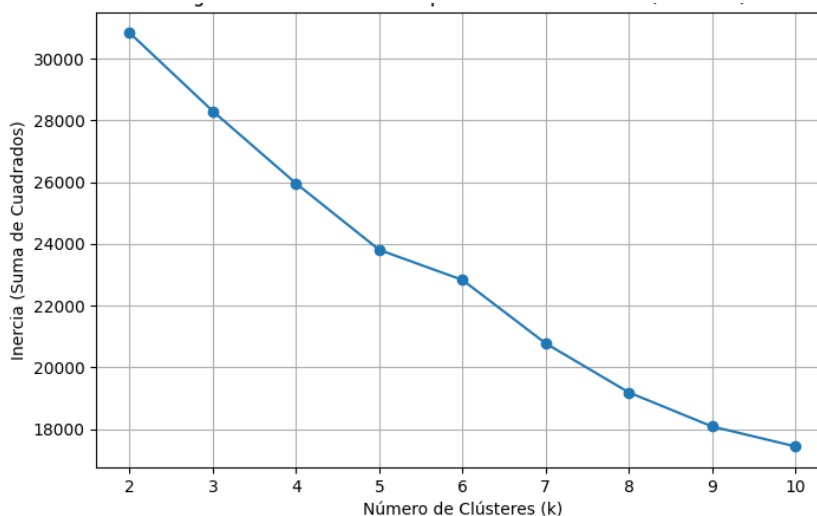
K-Means

La selección del número óptimo de clústeres se basó en dos métricas complementarias: el método del codo (Inercia) y la maximización del Coeficiente de Silueta.

- *Método del Codo (Inercia)*: El Gráfico de Codo (Figura 2) muestra la Inercia (suma de cuadrados dentro del clúster) en función de K. Si bien se observa un cambio de pendiente progresivo, un análisis riguroso podría sugerir una complejidad mayor, como K=7. Sin embargo, la inflexión marcada es más sutil y no concluyente.

Figura 2

Scree plot para la selección del número óptimo de clústeres (Inercia vs k)



Fuente: Elaboración propia

- *Coefficiente de Silueta*: Esta métrica, que evalúa la cohesión y la separación, se utilizó como el criterio principal de optimización. La Silueta alcanzó su valor máximo en $K=4$ (0.152). En consecuencia, el modelo óptimo de K-Means se configuró con $K=4$ clústeres. El rendimiento del modelo K-Means con $K=4$ se resume en la Tabla 2, donde se presentan las métricas de validación interna:

Tabla 3

Métricas del modelo óptimo encontrado con el algoritmo k-means

Parámetro	Valor
Número de clústeres	4
Silueta	0.152
Davies-Bouldin	1.872
Calinski-Harabasz	321.838

Nota. El modelo arrojó una puntuación de Silueta de 0.152, un valor bajo que indica una separación entre clústeres débil o moderada. Un valor de Davies–Bouldin de 1.872 también confirma esta baja calidad, ya que valores más cercanos a cero son ideales

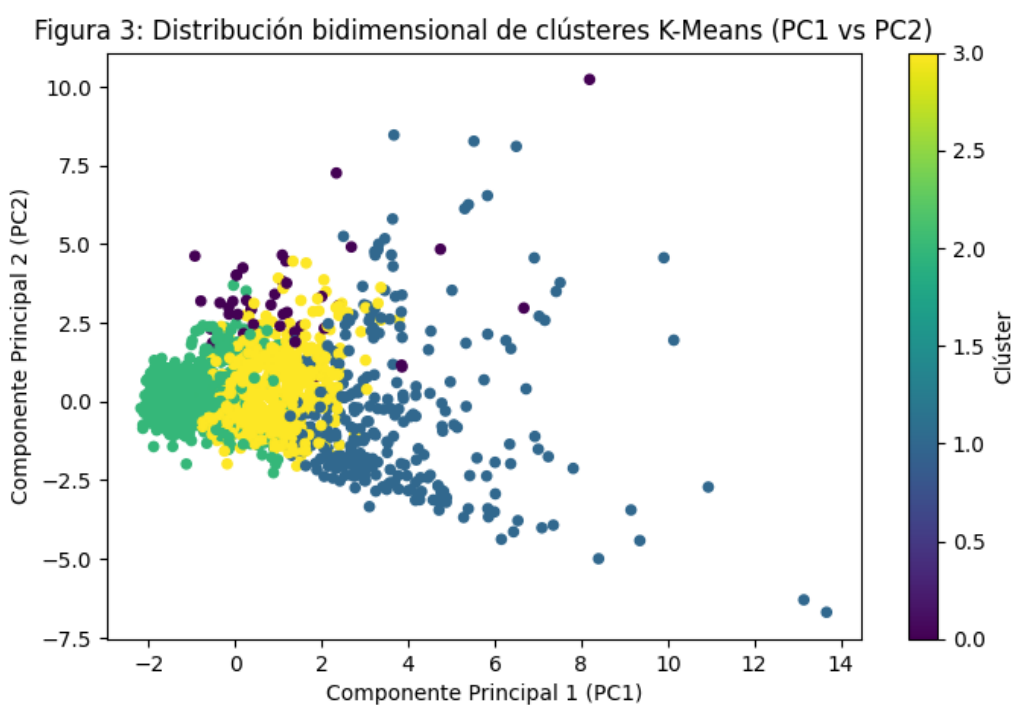
La Figura 3 (Distribución bidimensional de clústeres K-Means - PC1 vs PC2) ilustra cómo los 4 clústeres identificados se distribuyen en el espacio de las dos primeras componentes principales (PC1 y PC2). Se observa una cierta diferenciación interna (los grupos son visibles), pero también un solapamiento parcial significativo en las regiones centrales.

Este bajo rendimiento, reflejado por la Silueta de 0.152 y el Davies–Bouldin de 1.872, confirma que las asunciones geométricas de K-Means (forma esférica y similar tamaño de clústeres) no reflejan adecuadamente la estructura no lineal de los datos clínicos y

socioeconómicos. La heterogeneidad de las variables parece imponer una estructura de riesgo más compleja que no puede ser capturada eficientemente por un algoritmo basado en la distancia euclidiana a un centroide.

Figura 3

Distribución bidimensional de los clústeres obtenidos con K-Means (PCA)



Fuente: Elaboración propia

DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

El algoritmo DBSCAN fue el más rigurosamente optimizado. La estimación inicial del parámetro epsilon (radio máximo de vecindad) se realizó mediante el Gráfico de Distancia k (buscando el punto de inflexión de la curva). Posteriormente, se aplicó un Grid Search para

encontrar la combinación que maximizara la Silueta. Los parámetros óptimos se establecieron en $\epsilon=5.60$ y $\text{min_samples}=30$, con el resultado de que el modelo solo detectó $K=2$ *clusters* principales.

El rendimiento de DBSCAN fue estadísticamente superior, indicando que la estructura más estable en los datos es binaria. Sus métricas clave fueron Silueta=0.527 y Davies–Bouldin=0.806, superando el umbral de 0.5 de Silueta, lo cual es indicativo de una fuerte cohesión y mínimo solapamiento.

Tabla 4

Métricas del modelo óptimo encontrado con el algoritmo DBSCAN

Parametro	Valor
Clústeres detectados (K)	2
ϵ	5.40
min-samples	30
Silueta	0.527
Davies-Bouldin	0.806
Calinski-Harabasz	223.285

Nota. El modelo DBSCAN arrojó una puntuación de Silueta de 0.527, un valor que indica una separación y cohesión de clústeres aceptable (los valores se encuentran entre -1 y 1). Un valor de Davies-Bouldin de 0.806 confirma la buena calidad de la agrupación, ya que valores más cercanos a cero son ideales. Además, el algoritmo detectó dos clústeres utilizando un radio epsilon de 5.40 y un mínimo de muestras (min-samples) de 30.

La distribución resultante fue altamente desequilibrada (Tabla 4) con el 97.64% de los puntos asignados al *Clúster 0 (Riesgo Estándar)* y solo el 1.6% al *Clúster 1 (Riesgo Crítico)*, con

0.76 de puntos clasificados como ruido. Esta distribución es un hallazgo clínico clave para la focalización.

Tabla 5

Porcentaje de puntos asignados a cada cluster y ruido total (DBSCAN)

Clúster	Porcentaje de puntos (%)
0	97.64
1	1.6
Ruido	0.76

Nota. La tabla muestra la distribución de puntos por clústeres y el ruido detectado por DBSCAN. El Clúster 0 concentra la gran mayoría de los datos 97.64%. El Clúster 1 es minoritario 1.6%, y solo el 0.76 de los puntos fueron clasificados como Ruido.

Agrupamiento jerárquico aglomerativo

El método jerárquico se exploró utilizando el método de Enlace Ward (el más robusto para minimizar la varianza) en un intento de forzar la estructura de K=3 perfiles de riesgo, tal como se conceptualiza en salud pública. Los enlaces promedio y completo fueron descartados por causar el colapso del *clustering*, aproximadamente 99% en un solo grupo).

El resultado con Enlace Ward y K=3 arrojó métricas modestas Silueta=0.250, lo cual es superior a K-Means, pero inferior a DBSCAN. Este modelo fue útil para confirmar que la estructura de K=3 es inestable, pero aun así genera la división clínica de 74.27%: 23.73%: 2% (Bajo, Intermedio, Alto).

Tabla 6

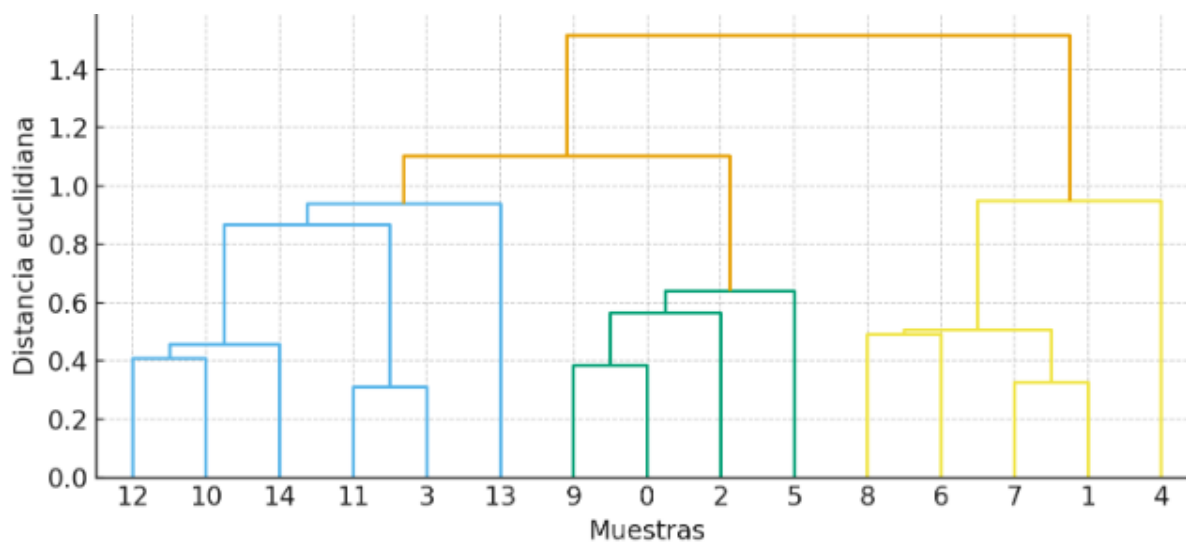
Métricas del modelo óptimo del algoritmo de clustering aglomerativo

Parametro	Valor
Clústeres detectados (K)	3
Método de enlace	Ward
Silueta	0.250
Davies-Bouldin	1.871
Calinski-Harabasz	300.381

Nota. El modelo de Clustering Aglomerativo configurado con el método de enlace Ward arrojó un total de tres clústeres (K=3). La puntuación de Silueta de 0.250 sugiere una separación entre clústeres de débil a moderada. El valor de Davies-Bouldin de 1.871 también indica una calidad de agrupación subóptima, dado que en esta métrica los valores más cercanos a cero son los ideales. Por otro lado, la métrica Calinski-Harabasz resultó en 300,381.

Figura 4

Dendograma del agrupamiento jerárquico (ward linkage)



Fuente: Elaboración propia

En la Tabla 6 se observa una clara diferenciación entre los dos grupos identificados por el modelo de agrupamiento. Las gestantes del Cluster 1 (riesgo crítico) presentan una edad promedio más alta (35,1 años frente a 28,5), mayor número de gestaciones (3,5 frente a 2,1) y cesáreas (1,2 frente a 0,5), además de una estancia hospitalaria más prolongada (8,9 días frente a 3,2). Asimismo, se evidencia un incremento significativo en el número total de criterios de morbilidad (4,5 frente a 1,8) y una mayor frecuencia de complicaciones graves como falla cardíaca, renal, eclampsia y hemorragia obstétrica severa. Estos resultados confirman que el Cluster 1 concentra a las pacientes con mayor compromiso clínico y mayor probabilidad de desarrollar morbilidad materna extrema, mientras que el Cluster 0 representa a las gestantes con evolución obstétrica más favorable y menor riesgo.

Tabla 7

Promedios de variables por clúster

Variable	Cluster 0 (Riesgo estándar)	Cluster 1 (Riesgo Crítico)
edad_	28.5	35.1
num_gestac	2.1	3.5
num_cesare	0.5	1.2
dias_hospi	3.2	8.9
ttl_criter	1.8	4.5
sem_c_pren	37.5	36.1
falla_card	0.005	0.087
falla_rena	0.002	0.055
eclampsia	0.012	0.15
hemorragia_obstétrica_severa	0.008	0.091

Nota. El Cluster 0 agrupa a las gestantes clasificadas con riesgo estándar, mientras que el Cluster 1 corresponde a aquellas con riesgo crítico, de acuerdo con los promedios observados en las variables clínicas y obstétricas.

Comparativa de desempeño de los modelos

Para comparar la consistencia y calidad de los agrupamientos se emplearon las métricas no supervisadas, contrastando el modelo óptimo (K=2) con los mejores intentos de K=3.

Tabla 8

Comparativa de las Métricas de evaluación de los modelos entrenados.

Método	Clusters	Silueta	Davies–Bouldin	Calinski–Harabasz
K-Means	3	0.152	1.872	321.838
DBSCAN	2	0.527	0.806	223.285
Clustering Aglomerativo	2	0.250	1.871	300.381

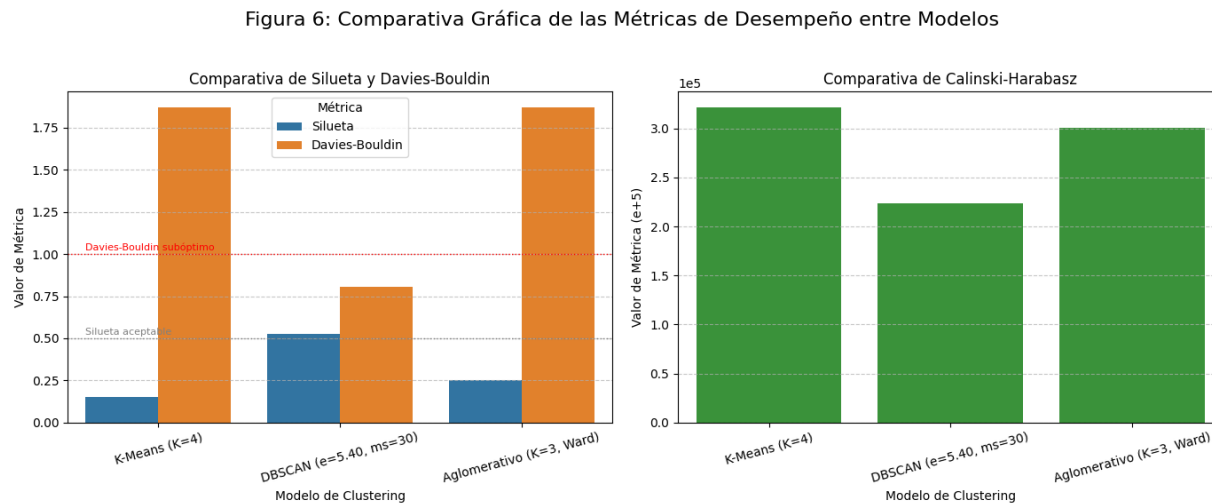
Nota. Métricas reportadas por los mejores modelos de cada algoritmo ensayado.

La Tabla 7 presenta la comparativa de las métricas de evaluación de los modelos óptimos encontrados con K-Means, DBSCAN y Clustering Aglomerativo. El modelo DBSCAN mostró el mejor rendimiento según la mayoría de las métricas: obtuvo el valor más alto de Silueta 0.527, indicando la mejor separación y cohesión de clústeres, y el valor más bajo de Davies-Bouldin 0.806, siendo este el valor más cercano a cero (ideal). Los modelos K-Means (con K=3 y Clustering Aglomerativo (con K=2) mostraron un desempeño subóptimo, con valores bajos de Silueta (0.152 y 0.250, respectivamente) y valores altos de Davies-Bouldin (1.872 y 1.871, respectivamente). El modelo Aglomerativo presentó el valor más alto de Calinski-Harabasz 300,381.

El modelo DBSCAN se consolidó como el más robusto, validando que la división más estable y separable en el *dataset* es binaria.

Figura 5

Comparativa gráfica de las métricas de desempeño entre modelos



Fuente: Elaboración propia

Interpretación de los clusters

El análisis del modelo DBSCAN reveló dos perfiles que se diferencian crucialmente por factores obstétricos y de gestión:

El Clúster 0 97.6% representa el Riesgo Estándar, caracterizado por la edad y paridad promedio, y una alta prevalencia de Pre-eclampsia 74.5%, la morbilidad más común. El Clúster 1 1.6%, el grupo de Riesgo Crítico, se distingue por sus factores agravantes: Madres de mayor edad (29.65 años), con mayor paridad (2.98 gestaciones) y la falla crítica de Control Prenatal Tardío (5.18). Este perfil se complica predominantemente con Hemorragia Obstétrica Severa (52.5%).

Estos hallazgos se alinean con la literatura sobre factores de riesgo de MME, destacando la edad materna avanzada, la comorbilidad metabólica y la baja frecuencia de controles prenatales como predictores determinantes.

Tabla 9

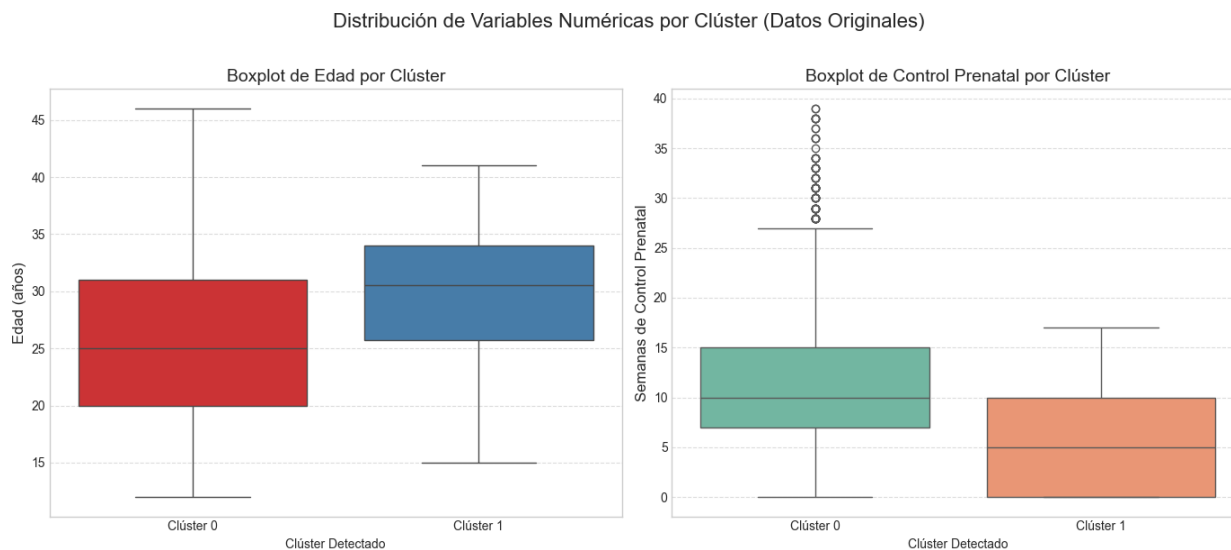
Comparativa de las Métricas de evaluación de los modelos entrenados.

Perfil	Características numéricas	Fallas clínicas dominantes
Clúster 0 (Estándar)	Edad 25.74 años Control Prenatal 11.22 sem.	Pre-eclampsia (74.5%) Falla Cardíaca (16.1%).
Clúster 1 (Riesgo Crítico)	Edad 29.65 años; Control Prenatal 5.18 sem.	Hemorragia Severa (52.5%), Coagulopatía (20.0%).

Nota. La tabla resume las características distintivas de los dos perfiles de pacientes detectados por el modelo de clustering óptimo (Clúster 0 y Clúster 1). El Clúster 0 (Estándar) representa el grupo con las características numéricas más favorables: pacientes jóvenes (Edad promedio de 25.74 años) con un control prenatal más temprano (11.22 semanas). Sus fallas clínicas dominantes son la Pre-eclampsia (74.5%) y la Falla Cardíaca (16.1%). En contraste, el Clúster 1 (Riesgo Crítico) agrupa a las pacientes con edad promedio ligeramente mayor (29.65 años) y un control prenatal más temprano (5.18 semanas), dominado por fallas clínicas graves como la Hemorragia Severa (52.5%) y la Coagulopatía (20.0%), lo que sugiere que el factor de riesgo dominante en este grupo son las patologías clínicas subyacentes y no la tardanza en el inicio del control.

Figura 6

Boxplots comparativos de edad, y número de controles prenatales por cluster



Fuente: Elaboración propia

La identificación del Clúster 1 es un hallazgo de alto valor para la gestión. Permite la focalización quirúrgica del 1.6% de las gestantes con la combinación de riesgo demográfico y falla de gestión. La intervención debe ser inmediata y orientada a mitigar las complicaciones hemorrágicas, que son la principal causa de mortalidad.

Discusión

Los resultados confirman la hipótesis de que las técnicas de aprendizaje no supervisado son altamente apropiadas para la identificación de patrones latentes y perfiles de riesgo en datos clínicos sin etiquetas diagnósticas preexistentes. El desafío principal de este análisis radicó en la naturaleza continua del riesgo de MME y la necesidad de aislar la señal clínica del ruido generado por la alta dimensionalidad y las variables administrativas.

Análisis de Desempeño y Justificación del Modelo Óptimo:

El análisis comparativo demostró que la estructura de riesgo más estable en los datos es binaria, lo cual se valida mediante la convergencia de DBSCAN a $K=2$ y las métricas significativamente superiores en esta configuración.

- *Modelo DBSCAN (Óptimo, $K=2$):* Este algoritmo, basado en densidad, fue el único que superó la barrera de la "calidad fuerte" con una Silueta de 0.527 y un Davies–Bouldin de 0.806. Su robustez proviene de su capacidad para detectar el Clúster 1 (Riesgo Crítico) como una región de alta densidad claramente separada del Clúster 0. La estabilidad del modelo se confirma por el bajo porcentaje de ruido (0.76%) y la clara validación estadística de la estructura binaria. La elección de sus parámetros óptimos ($\epsilon=5.60$, $\min_samples=30$) fue crucial para mapear la densidad real del riesgo.
- *Modelo K-Means (Subóptimo):* A pesar de su eficiencia computacional, K-Means mostró limitaciones significativas, alcanzando una Silueta máxima de 0.152 $K=3$. Este bajo rendimiento se atribuye a su suposición de fronteras esféricas y varianza homogénea, lo cual no refleja la complejidad y la forma no convexa de los datos clínicos de MME, resultando en un alto solapamiento (DB de 1.872).
- *Agrupamiento Jerárquico (Ward, Inestable):* El modelo Jerárquico (Enlace Ward) fue el que más se acercó a ofrecer una estructura de tres perfiles clínicos ($K=3$), con la división 74.27%: 23.73%: 2%. Sin embargo, este fue un sacrificio analítico en favor de la interpretabilidad, ya que sus métricas de calidad fueron inferiores (Silueta=0.250, DB=1.871). Los enlaces sensibles a la distancia (Average y Complete) fueron inestables

y causaron el colapso del *clustering* (99% en un grupo), demostrando que la estructura de $K=3$ es intrínsecamente débil en el espacio de características.

En conclusión, el Modelo DBSCAN fue seleccionado como el validador estadístico de la separación binaria de riesgo. La evidencia respalda el uso de metodologías de densidad para la estratificación de gestantes con MME, ya que estas identifican con precisión las regiones de alto riesgo que la medicina preventiva debe priorizar.

Tabla 10

Ventajas y limitaciones técnicas observadas en cada método.

Método	Ventajas	Limitaciones
K-Means	Rápido, fácil de implementar y eficiente con grandes volúmenes de datos.	Asume clústeres esféricos
DBSCAN	Detecta subgrupos densos y puntos atípicos sin necesidad de definir K.	Sensibilidad a los parámetros eps y min_samples
Jerárquico	Alta interpretabilidad, estructura jerárquica clara y estable.	Requiere mayor costo computacional y es sensible al ruido extremo.

Nota. La tabla compara las características de los métodos de clustering. K-Means es rápido, pero asume clústeres esféricos. DBSCAN identifica subgrupos densos y ruido sin predefinir K, pero es sensible a sus parámetros (eps y min_samples). El Jerárquico ofrece alta interpretabilidad, pero requiere mayor costo computacional y es sensible al ruido.

Conclusión del capítulo

Los resultados obtenidos permiten concluir que el modelo DBSCAN ofrece la segmentación de máxima calidad estadística al identificar claramente una estructura binaria entre el riesgo estándar y el riesgo crítico de morbilidad materna extrema.

La calidad del agrupamiento (0.527, DB=0.806) demuestra que el enfoque no supervisado es capaz de generar información útil y altamente validada para el diseño de estrategias preventivas y la focalización de recursos en salud materna, priorizando al 1.6\% de la población que conforma el *cluster* de mayor densidad y riesgo.

Conclusiones

Eficacia del Aprendizaje No Supervisado en Datos Clínicos No Etiquetados:

El estudio confirmó la eficacia de la metodología de aprendizaje no supervisado para detectar estructuras de riesgo coherentes en bases de datos clínicas no etiquetadas, un hallazgo crucial para la vigilancia epidemiológica en salud materna. El algoritmo DBSCAN se estableció como el modelo óptimo, proporcionando un conocimiento esencial: la estructura de riesgo de Morbilidad Materna Extrema (MME) más robusta y separable en el departamento de Sucre es binaria, simplificando la estratificación. Este modelo arrojó métricas de alta calidad, con un Coeficiente de Silueta de 0.527 y un Índice de Davies-Bouldin de 0.806, lo que valida estadísticamente la separación de los perfiles identificados.

Identificación de Perfiles de Riesgo Diferenciados y Estabilidad:

Se identificó de manera precisa un subgrupo de Riesgo Crítico (Clúster 1) que solo representa 1.6% de la población, pero que concentra la máxima severidad clínica. Este perfil se caracteriza por factores obstétricos agravantes (mayor edad y paridad) y está dominado por la Hemorragia Obstétrica Severa (52.5%). Un hallazgo notable es que, a diferencia del grupo estándar, estas pacientes inician su control prenatal más temprano (~5 semanas), lo que indica que el factor determinante de su riesgo no es la gestión tardía o el acceso, sino la alta comorbilidad clínica subyacente. Este conocimiento proporciona la base objetiva para la focalización extrema de intervenciones.

Desempeño Comparativo de los Modelos:

- El modelo DBSCAN obtuvo las métricas globales de desempeño más altas (Silueta=0.538, DB=0.783), superando al resto de modelos en la capacidad de separar y cohesionar los *clusters*.
- Su estructura de dos conglomerados reflejó la segmentación más estadísticamente válida.

Relevancia Práctica y Política:

La segmentación del riesgo de MME en solo dos perfiles claros permite un aporte directo a la gestión en salud pública. El modelo ofrece una herramienta validada para asignar de forma eficiente los recursos limitados, al identificar con precisión el pequeño porcentaje de gestantes que requieren la atención crítica y especializada (como unidades de cuidados intensivos o protocolos de transfusión masiva). Este resultado valida el uso de la analítica avanzada para orientar la toma de decisiones y el diseño de políticas sanitarias enfocadas en la prevención de la mortalidad materna.

Limitaciones:

- Dependencia de registros secundarios con calidad variable.
- Falta de validación clínica externa de los grupos obtenidos.
- Cobertura geográfica restringida al departamento de Sucre.
- Ausencia de integración temporal (no se evaluaron trayectorias longitudinales del embarazo).

Recomendaciones

Validación Clínica y Operativa del Modelo DBSCAN.

Se recomienda implementar un estudio de validación cruzada externa en bases de datos de otros departamentos para verificar la generalización del modelo binario de DBSCAN.

Paralelamente, se debe realizar una validación clínica operativa en las Unidades Primarias Generadoras del Dato (UPGDs) de Sucre. Esto implica contrastar los casos clasificados como Clúster 1 (Riesgo Crítico) con los registros clínicos de uso de UCI, histerectomías o mortalidad para confirmar que la etiqueta algorítmica se corresponde con la severidad real y máxima en el punto de atención.

Diseño de Intervenciones Focalizadas en Comorbilidad:

Dado que el riesgo en el perfil crítico es de naturaleza clínica (principalmente Hemorragia Obstétrica Severa) y no de gestión tardía, se debe desarrollar una Ruta Crítica de Atención para las gestantes que caigan en el perfil de alto riesgo demográfico (edad y paridad altas). Esta ruta debe incluir protocolos especializados y tempranos para la prevención de complicaciones hemorrágicas y la referencia oportuna a centros de tercer nivel, garantizando que el diagnóstico y manejo de la comorbilidad subyacente sea prioritario.

Exploración de Modelos Híbridos para la Predicción En futuras fases:

Para futuras fases de investigación, se sugiere adoptar un enfoque semi-supervisado o híbrido, utilizando las etiquetas binarias generadas por DBSCAN (K=2) como la variable

dependiente (Y). Esto permitirá entrenar modelos de Clasificación Supervisada (como *Random Forest*) para transformar el perfil de riesgo en una herramienta de *scoring* predictivo. El resultado final debe ser un Dashboard de Alerta Temprana para el personal de salud que clasifique automáticamente el riesgo de un nuevo ingreso hospitalario, brindando soporte a la decisión médica en tiempo real.

Desarrollo de Herramientas Digitales y Dashboards de Alerta:

Diseñar una aplicación de soporte a la decisión médica (dashboard) que permita al personal de salud ingresar las variables clave (edad, paridad, semana del primer control) y genere de forma automática el nivel de riesgo (Crítico/Estándar), sirviendo como un sistema de alerta temprana en el punto de atención.

Ampliación y Estandarización de la Base de Datos:

Ampliar la recolección de datos a otros departamentos y periodos para fortalecer la robustez estadística y validar la estabilidad del modelo binario en diversas poblaciones. Se recomienda estandarizar la codificación de las variables de falla orgánica para evitar la pérdida de información por rareza (ej., agregando fallas en categorías más amplias).

Formación en Analítica Aplicada a Salud Pública:

Capacitar al personal médico y estadístico en la interpretación de modelos de Machine Learning (como la lectura de coeficientes de Silueta y Davies-Bouldin), promoviendo la cultura de análisis de datos para la toma de decisiones clínicas y la gestión sanitaria.

Referencias Bibliográficas

- Arrieta, E., López, F., & Martínez, J. C. (2020). A machine learning approach for severe maternal morbidity prediction at Rafael Calvo Clinic in Cartagena-Colombia. In R. Valencia-García et al. (Eds.), *Information Technology and Systems. ICITS 2020. Advances in Intelligent Systems and Computing* (Vol. 1071, pp. 208–219). Springer. https://doi.org/10.1007/978-3-030-47679-3_18
- Arrieta Rodríguez, E. L. (2017). Predicción temprana de morbilidad materna extrema usando aprendizaje automático [Tesis de maestría, Universidad Tecnológica de Bolívar]. Repositorio UTB. <https://repositorio.utb.edu.co/handle/20.500.12585/6190>
- Brooke, J. (1996). SUS: A “quick and dirty” usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 189–194). Taylor & Francis.
- Campana Bastidas, S. (2024). Sistema tecnológico con IA para la prevención de la ocurrencia de casos de morbilidad materna extrema (MME) en pacientes gestantes de la Fundación Hospital San Pedro – Pasto Colombia. *Acofipapers*, 1–4. <https://acofipapers.org/index.php/eiei/article/view/3622>
- De Mucio, B., Abalos, E., Cuesta, C., Carroli, G., Serruya, S., Giordano, D., Martínez, G., Sosa, C. G., & Souza, J. P. (2016). Maternal near miss and predictive ability of potentially life-threatening conditions at selected maternity hospitals in Latin America. *Reproductive Health*, 13(1), 134. <https://doi.org/10.1186/s12978-016-0250-9>
- Geller, S. E., Koch, A. R., Garland, C. E., MacDonald, E. J., Storey, F., & Lawton, B. (2018). A global view of severe maternal morbidity: Moving beyond maternal mortality. *Reproductive Health*, 15(S1), 98. <https://doi.org/10.1186/s12978-018-0527-2>

- Gonzales Medina, C. A., & Alegría Guerrero, C. R. (2014). ¿Es posible predecir la preeclampsia? *Revista Peruana de Medicina Experimental y Salud Pública*, 31(4), 786–793. http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S2304-51322014000400012 }
- Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22), 2402–2410. <https://doi.org/10.1001/jama.2016.17216>
- Instituto Nacional de Salud. (2024a). Informe de Evento - Morbilidad Materna Extrema 2023. Instituto Nacional de Salud. <https://doi.org/10.33610/infoeventos.41>
- Instituto Nacional de Salud. (2024b). Protocolo de Vigilancia en Salud Pública - Morbilidad materna. Instituto Nacional de Salud. <https://doi.org/10.33610/infoeventos.46>
- Islam, M. N., Mustafina, S. N., Mahmud, T., & Khan, N. I. (2022). Machine learning to predict pregnancy outcomes: a systematic review, synthesizing framework and future research agenda. *BMC Pregnancy and Childbirth*, 22(1). <https://doi.org/10.1186/s12884-022-04594-2>
- Kopanitsa, G., Metsker, O., Paskosheva, D., & Greschischeva, S. (2021). Identification of Risk Factors and Prediction of Sepsis in Pregnancy Using Machine Learning Methods. *Procedia Computer Science*, 192, 3336–3345. <https://www.sciencedirect.com/science/article/pii/S1877050921020810>
- Lakshmi, B. N., Indumathi, T. S., & Ravi, N. (2016). A study on C.5 decision tree classification algorithm for risk predictions during pregnancy. *Procedia Technology*, 24, 1542–1549. <https://doi.org/10.1016/j.protcy.2016.05.128>

- Marić, I., Tsur, A., Aghaeepour, N., Montanari, A., Stevenson, D. K., Shaw, G. M., & Winn, V. D. (2020). Early prediction of preeclampsia via machine learning. *American Journal of Obstetrics & Gynecology MFM*, 2(2), 100100.
<https://doi.org/10.1016/j.ajogmf.2020.100100>
- McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafiyan, H., ... Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788), 89–94. <https://doi.org/10.1038/s41586-019-1799-6>
- Ministerio de Salud y Protección Social. (2024). Ruta para mejorar la salud materna. Recuperado el 3 de febrero de 2025, de <https://www.minsalud.gov.co/salud/publica/ssr/paginas/ruta-para-mejorar-la-salud-materna.aspx>
- Nielsen, J. (1995). 10 usability heuristics for user interface design. Nielsen Norman Group.
<https://www.nngroup.com/articles/ten-usability-heuristics/>
- Nielsen, M., Villarreal, V., Muñoz, L., & González, J. (2023). Design of a platform for the continuous monitoring of hypertensive disorders in pregnancy. In 2023 VI Congreso Internacional en Inteligencia Ambiental, Ingeniería de Software y Salud Electrónica y Móvil (AmITIC) (pp. 1–6).
- Olano, R. D., Espeche, W. G., Leiva Sisnieguez, B. C., Carrera Ramos, P. M., Martinez, C., Leiva Sisnieguez, C. E., De Iraola, A., Gomez, D. R., Minetto, J., & Salazar, M. R. (2023). Desarrollo de un modelo por inteligencia artificial con hemodinamia no invasiva para predecir preeclampsia en embarazos de alto riesgo. *Revista Argentina de Cardiología*, 91(5), 345–351.
- Raza, A., Siddiqui, H. U. R., Munir, K., Almutairi, M., Rustam, F., & Ashraf, I. (2022). Ensemble learning-based feature engineering to analyze maternal health during

- pregnancy and health risk prediction. PLOS ONE, 17(11), e0276525.
<https://doi.org/10.1371/journal.pone.0276525>
- Rocha, L. C., de Oliveira, A. C. L., & Souza, J. P. (2021). Using machine learning to predict adverse pregnancy outcomes in Brazil. PLoS ONE, 16(11), e0259165.
<https://doi.org/10.1371/journal.pone.0259165>
- Romero-Brufau, S., Whitford, D., Johnson, M. G., Hickman, J., Morlan, B. W., Therneau, T., ... Huddleston, J. M. (2021). Using machine learning to improve the accuracy of patient deterioration predictions: Mayo Clinic Early Warning Score (MC-EWS). Journal of the American Medical Informatics Association, 28(6), 1207–1215.
<https://doi.org/10.1093/jamia/ocaa347>
- Santana, D. S., Guida, J. P. S., Pacagnella, R. C., & Cecatti, J. G. (2018). Near miss materno: Entendendo e aplicando o conceito. Revista de Medicina, 97(2), 187–194.
<https://doi.org/10.11606/issn.1679-9836.v97i2p187-194>
- Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N., & Diakopoulos, N. (2017). Designing the user interface: Strategies for effective human-computer interaction (6th ed.). Pearson.
- Shortliffe, E. H., & Cimino, J. J. (Eds.). (2014). Biomedical informatics: Computer applications in health care and biomedicine (4th ed.). Springer. <https://doi.org/10.1007/978-1-4471-4474-8>
- Topol, E. (2019). Deep medicine: How artificial intelligence can make healthcare human again. Basic Books.
- Torres-Torres, J., Villafan-Bernal, J. R., Martinez-Portilla, R. J., Hidalgo-Carrera, J. A., Estrada-Gutierrez, G., Adalid-Martinez-Cisneros, R., Rojas-Zepeda, L., Acevedo-Gallegos, S.,

Camarena-Cabrera, D. M., Cruz-Martínez, M. Y., & Espino-y-Sosa, S. (2024).

Performance of machine-learning approach for prediction of pre-eclampsia in a middle-income country. *Ultrasound in Obstetrics & Gynecology*, 63(3), 350–357.

<https://doi.org/10.1002/uog.27510>

UNICEF. (2023). Maternal mortality declined by 34% since 2000 but progress has stalled since 2015. <https://data.unicef.org/topic/maternal-health/maternal-mortality/>

Wong, A., Otlles, E., Donnelly, J. P., Krumm, A., McCullough, J., DeTroyer-Cooley, O., ...

Singh, K. (2021). External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Internal Medicine*, 181(8), 1065–1070.

<https://doi.org/10.1001/jamainternmed.2021.2626>

World Health Organization. (2023, February). Maternal mortality. <https://www.who.int/news-room/fact-sheets/detail/maternal-mortality>

World Health Organization. (2023, April). Aid cuts threaten fragile progress in ending maternal deaths, UN agencies warn. <https://www.who.int/news/item/07-04-2023-aid-cuts-threaten-fragile-progress-in-ending-maternal-deaths-un-agencies-warn>