



ALGORITMO DE UNA RED NEURONAL PROFUNDA PARA LA PREDICCIÓN
DE RIESGOS DE MORBILIDAD MATERNA EXTREMA (MME) EN LA E.S.E
CLÍNICA DE MATERNIDAD RAFAEL CALVO

PRESENTADO POR:
DAIRO ELIECER MONSALVE ESPINOSA
KEVIN ANDRES RODRIGUEZ OTERO

UNIVERSIDAD DEL SINÚ ELIAS BECHARA ZAINUM
FACULTAD DE CIENCIAS EXACTAS E INGENIERIA
ESCUELA DE INGENIERÍA DE SISTEMAS
CARTAGENA, BOLÍVAR
NOVIEMBRE 2019



UNIVERSIDAD DEL SINÚ
Elías Bechara Zainúm
Seccional Cartagena

ALGORITMO DE UNA RED NEURONAL PROFUNDA PARA LA PREDICCIÓN
DE RIESGOS DE MORBILIDAD MATERNA EXTREMA (MME) EN LA E.S.E
CLÍNICA DE MATERNIDAD RAFAEL CALVO.

Trabajo de grado presentado para como requisito para optar el título de
INGENIERIO DE SISTEMAS

Asesor Disciplinar
EUGENIA ARRIETA RODRIGUEZ

Asesor Metodológico:
EUGENIA ARRIETA RODRIGUEZ

UNIVERSIDAD DEL SINÚ ELIAS BECHARA ZAINUM
FACULTAD DE CIENCIAS EXACTAS E INGENIERIA
ESCUELA DE INGENIERÍA DE SISTEMAS
CARTAGENA – BOLÍVAR
NOVIEMBRE-2019

ACTA DE CALIFICACIÓN Y APROBACIÓN

Nota de aceptación

Director de Escuela

Director de Investigaciones

Firma del jurado

Firma del jurado

Cartagena - (20, 11,2019)

TABLA DE CONTENIDO

RESUMEN.....	10
INTRODUCCIÓN	11
1. DISEÑO METODOLÓGICO	13
1.1. Planteamiento Del Problema	13
1.1.1. Descripción del problema	13
1.1.2. Formulación de problema.....	14
1.1.3. Justificación del problema	15
1.2. Objetivos.....	16
General.....	16
Específicos	16
1.3. Alcance del proyecto.....	17
1.4. Marco de referencia	18
1.4.1. Estado del arte	18
1.4.2. Marco teórico	19
1.4.3. Marco conceptual	29
1.4.4. Consideraciones éticas y marco legal	33
1.5. Metodología	36
1.5.1. Línea de investigación.....	36
1.5.2. Tipo de investigación.....	37
1.5.3. Descripción de la metodología	37
2. CONOCIMIENTO DEL NEGOCIO	40
2.1. Consulta preconcepcional.....	40
2.2. Proceso de atención	41
2.2.1. Protocolo de Morbilidad Materna Extrema código INS: 549.....	42
2.3. Consulta prenatal.....	44
2.4. Criterios de inclusión.....	45
3. ADQUISICIÓN Y COMPRESIÓN DE LOS DATOS	46

3.1.	Calidad de los datos	46
3.1.1.	Obtención de los datos.....	46
3.1.2.	Análisis de las variables de riesgo	54
3.1.3.	Construcción del conjunto de datos	59
4.	MODELADO DE LA RED NEURONAL	61
4.1.	Técnicas de filtrado de variables	62
4.2.	Normalización de los datos.....	63
4.3.	División de los datos, entrenamiento y prueba	63
4.4.	Balanceo de los datos.....	64
4.5.	Construcción del algoritmo de predicción	66
4.6.	Entrenamiento del modelo	68
4.7.	Evaluación del modelo.....	69
5.	RESULTADOS	75
	CONCLUSIONES Y RECOMENDACIONES	80
	BIBLIOGRAFÍA.....	85
	CRONOGRAMA	87

TABLA DE ILUSTRACIONES

	Pág.
Ilustración 1 Red neuronal organizada en capas. - Fuente propia.....	24
Ilustración 2 Distribución de pesos en una red neuronal - Fuente Propia.....	24
Ilustración 3 Función para reducir la función de coste y mejorar predicciones	26
Ilustración 4 Ciclo de vida metodología TDSP de Microsoft [15].....	38
Ilustración 5 Protocolo de Morbilidad Materna Extrema Cod INS 549 – tomado de https://www.ins.gov.co/Paginas/Inicio.aspx	42
Ilustración 6 Protocolo de Morbilidad Materna Extrema Cod INS 549 –tomado de https://www.ins.gov.co/Paginas/Inicio.aspx	42
Ilustración 7 Protocolo de Morbilidad Materna Extrema Cod INS 549 - 3 - tomado de https://www.ins.gov.co/Paginas/Inicio.aspx	43
Ilustración 8 Código SQL – Fuente propia	47
Ilustración 9 Código SQL – Fuente propia	47
Ilustración 10 Código SQL – Fuente propia	48
Ilustración 11 Código SQL – Fuente propia	49
Ilustración 12 Código SQL – Fuente propia	50
Ilustración 13 Código SQL – Fuente propia	50
Ilustración 14 Código SQL – Fuente propia	51
Ilustración 15 Código SQL – Fuente propia	51
Ilustración 16 Código SQL – Fuente propia	52
Ilustración 17 Código SQL – Fuente propia	52
Ilustración 18 Código SQL – Fuente propia	53
Ilustración 19 Código SQL – Fuente propia	53
Ilustración 20 Código SQL – Fuente propia	54
Ilustración 21 Código SQL – Fuente propia	54
Ilustración 22 Valores nulos en la columna MESES_SEMANAS.....	56
Ilustración 23 Columna MESES_SEMANAS con datos nulos	56
Ilustración 24 Cantidad de datos nulos	57
Ilustración 25 Conjunto de datos, sin la columna MESES_SEMANAS.....	57
Ilustración 26 Cantidad de 0 y 1, relación a la variable de salida.....	58
Ilustración 27 - 0 y 1 (entrenamiento)	58
Ilustración 28 Rango de edades relación al conjunto de datos.....	59
Ilustración 29 Conjunto de datos – fuente propia.....	60
Ilustración 30 Conjunto de datos – fuente propia.....	60
Ilustración 31 Conjunto de datos - fuente propia.....	60

Ilustración 32 Diseño del Algoritmo.....	61
Ilustración 33 Resultado de la técnica de filtrado RFE.....	62
Ilustración 34 Resultado de la técnica de filtrado.....	62
Ilustración 35 Resultado de la normalización.....	63
Ilustración 36 Split de entrenamiento y prueba - (Fuente propia)	64
Ilustración 37 Datos desbalanceados	64
Ilustración 38 Resultado de la clase 0 y 1 balanceadas	65
Ilustración 39 Resultado exactitud y perdida ; entrenamiento (train) vs validación (test).....	70
Ilustración 40 Resultado en la matriz de confusión.....	71
Ilustración 41 Curva ROC ; VP y FP	74
Ilustración 42 Construcción de la red neuronal profunda para la predicción de riesgos de MME	75
Ilustración 43 Resultado de las métricas	83

LISTA DE TABLAS

Pág.

Tabla 1 Criterios de inclusión de MME - Fuente propia	21
Tabla 2 Definición de criterios de inclusión de MME - Fuente propia	23
Tabla 3 Función de activación Relu	31
Tabla 4 Función de activación Softmax	31
Tabla 5 Cumplimiento de los objetivos con base a la metodología.....	39
Tabla 6 Criterios de inclusión de MME - Fuente propia	45
Tabla 7 Resultado de las cantidades de filas y columnas en la división del Dataset – Fuente propia.....	64
Tabla 8 Resultado de la técnica smote, balanceando la cantidad de 0 y 1	65
Tabla 9 Resultado con la cantidad de filas y columnas balanceadas en el Dataset - fuente propia	66
Tabla 10 Resultado de la visión de la variable dependiente MME.....	66
Tabla 11 Construcción de la red neuronal	67
Tabla 12 Parámetros para la compilación del modelo	67
Tabla 13 Parámetros para el entrenamiento del modelo	68
Tabla 14 Resultados del entrenamiento del modelo	68
Tabla 15 Resultado con base a la precisión y perdida.....	69
Tabla 16 Grafica del margen de error ; entrenamiento (train) vs validación (test) .	70
Tabla 17 Resultado de la evaluación del modelo - train vs test	71
Tabla 18 Resultado del bloque de entrenamiento del modelo	77
Tabla 19 Resultados de las métricas	78

LISTA DE ANEXOS

	Pág.
Anexo A. Cronograma	87

RESUMEN

Este trabajo presenta el desarrollo de un algoritmo de inteligencia artificial mediante la implementación de técnicas de aprendizaje automático supervisado, que permiten predecir el riesgo de Morbilidad Materna Extrema (MME) con un nivel de precisión del 66% y 66% de sensibilidad. Esta herramienta busca prevenir y mejorar la calidad vida de las mujeres que lleguen a presentar casos de MME, el cual está considerado por la Organización Mundial de la Salud (OMS) como un problema de salud pública tanto en Colombia como en el mundo ya que esta condición es altamente riesgosa tanto para la madre como para el bebé.

Para el desarrollo del algoritmo se contó con la aprobación de la E.S.E. Clínica de Maternidad Rafael Calvo (CMRC) de Cartagena de Indias, la cual proporciono un conjunto de datos de las gestantes atendidas en esa entidad, la cual brinda atención durante su proceso de embarazo, cabe resaltar que el manejo de la información es de carácter confidencial e investigativo y se encuentra anonimizada.

Al conjunto de datos (Dataset) se le aplicaron técnicas de filtrado de datos mediante SQL server para tener datos agrupados, clasificados y filtrados. Al final el Dataset quedara con datos relevantes y optimizados para el entrenamiento de la red neuronal que se va a desarrollar ya que se cuenta con muchas variables que son irrelevantes para el proyecto, posteriormente se realiza el filtrado de variables para el que se utiliza el 70% para entrenamiento y el otro 30% para pruebas.

INTRODUCCIÓN

La Organización Mundial de la Salud (OMS) define a la Morbilidad Materna Extrema (MME) como el caso de una mujer que sobrevive a una complicación de alto riesgo ocurrida durante el embarazo, parto o puerperio. Además, debe cumplir con al menos uno o más de los criterios de inclusión materna establecidos en Colombia como por ejemplo la eclampsia, preeclampsia severa, sepsis o infección sistémica severa entre otras [1].

En Colombia entre los años 2003 y 2006 se llevó a cabo la primera publicación sobre la revisión de casos de Morbilidad Materna Extrema, en el año 2007 se inicia por parte del Ministerio de Salud y Protección Social la estructuración de un Modelo de Vigilancia de la Morbilidad Materna Extrema. En el año 2010 se publica el documento conceptual con el primer análisis agregado de casos específicamente sobre la relación criterio/caso según causa principal, que para Colombia fue de 3,5, en tanto que en Latinoamérica fue de 2,9 [1].

En el año 2011 se inicia la construcción del primer protocolo de vigilancia en salud pública sobre la MME, el Instituto Nacional de Salud (INS) realiza una prueba piloto en diferentes entidades territoriales como el Valle del Cauca, Nariño, Bogotá, Caldas, Meta, Bolívar, Santander. Posterior a ello en el año 2012 el Instituto Nacional de Salud realiza la entrega del Protocolo de Vigilancia en Salud Pública de la (MME) a el resto de las entidades territoriales y se da inicio al proceso de vigilancia [1].

En Colombia y a nivel mundial, se ve identificado que las principales causas de MME se relacionan con las complicaciones obstétricas, fallas de disfunción orgánica y malos procedimientos clínicos. Por eso es importante conocer y aprender de todos estos factores de riesgos que con llevan a aun posible estado de muerte a una mujer en estado de embarazo, parto o puerperio [1].

Para evitar futuras muertes y poder controlar estos posibles riesgos, se lleva a cabo esta investigación con el uso principal de inteligencia Artificial que busca influir positivamente en la salud materna brindada a la sociedad colombiana, este caso en

particular se enfoca en la Clínica de Maternidad Rafael Calvo (CMRC), ubicada en la ciudad de Cartagena-Bolívar lugar donde se lleva a cabo la investigación de este proyecto el cual se basa en redes neuronales profundas (Deep Learning), con el objetivo de predecir una variable de salida Y con un alto, medio y bajo riesgo de MME esto con el fin de darle cumplimiento a los objetivos del milenio planteados en Colombia: Reducir la mortalidad infantil y mejorar los procesos de salud en la maternidad colombiana [1].

1. DISEÑO METODOLÓGICO

Este capítulo se enfoca en exponer el uso y el desarrollo de una red neuronal profunda aplicando librerías de Python construyendo así un algoritmo con el fin de brindar apoyo a los galenos, donde al ingresar las variables de riesgos obtendrá como resultado la predicción con un estado alto, medio, bajo riesgo donde deberá clasificarse como Morbilidad Materna Extrema o sana, es importante destacar que este tipo de herramientas no va a reemplazar la labor de un médico.

Adicionalmente es este capítulo se expone las causas y el porcentaje de mujeres gestantes que diariamente se ven afectadas con algunos de estas con las complicaciones obstétricas, fallas de disfunción orgánica y malos procedimientos clínicos, que conllevan a la Morbilidad Materna Extrema (MME), del mismo modo se plantean diversos estudios realizados por entidades educativas y prestadoras de servicios de la salud las cuales ayudan a ampliar el panorama y a entender más el tema tratado.

1.1. Planteamiento Del Problema

1.1.1. Descripción del problema

En Colombia y a nivel mundial la morbilidad materna extrema (MME) se ha convertido en un grave problema de salud pública afectando directamente a la población femenina cobrando diariamente muchas muertes. A pesar de los esfuerzos que realizan diariamente las entidades de salud, los profesionales de la salud y la misma comunidad, no se logra llevar a cabo lo planteado en los objetivos del milenio en Colombia: Reducir la mortalidad infantil y mejorar los procesos de salud en la maternidad colombiana [2].

Según estudios realizados por la Universidad Nacional (UN) a la población colombiana se demuestra que cerca de 600 mujeres gestantes mueren anualmente, haciendo de esto un alto índice de Mortalidad Materna en el país, de ellas el 16%

son adolescentes ya que presentan mayores dificultades para controlar su fecundidad [2].

En el estudio se logra demostrar que los departamentos con más alto índice de Mortalidad Materna son el Chocó, la Guajira, el Meta, el Guainía. Donde según especialistas una de las primeras causas de mortalidad es la hipertensión que ocupa un 33% [2] y otros factores de riesgo causantes de MME entre ellos enfermedades específicas como la eclampsia, preeclampsia, sepsis, hemorragias obstétricas, disfunciones orgánicas como falla renal, hepática, metabólica, cerebral, respiratoria y los malos procedimientos en atención a las pacientes, malos procedimientos quirúrgicos y malos procedimientos en UCI [3].

A pesar de los malos manejos, se resalta una gran oportunidad en el suceso de escuchar la vivencia o experiencia de aquella mujer que ha sobrevivido a un riesgo alto de MME ocurrida durante el embarazo, parto o puerperio, examinando claramente todos los riesgos o malos procedimientos en demoras de atención, el acceso a una salud de calidad en los servicios que brinde, lo que conlleva a ser un condicionante de riesgo para declararse en Morbilidad Materna Extrema. Estos elementos hacen parte de la determinación estructural social y de la intermedia, donde los servicios de salud son una expresión de malos manejos en muchos casos [3].

1.1.2. Formulación de problema

A partir de lo mencionado anteriormente surge una pregunta problema ¿Cómo desarrollar una red neuronal profunda que permita predecir los riesgos de MME en la etapa de posparto en la Clínica De Maternidad Rafael Calvo?

1.1.3. Justificación del problema

Durante los años 2012 – 2017 se llevó a cabo el plan de acción para acelerar la reducción de la Mortalidad Materna Extrema y futuros riesgos causantes de MME por lo cual se planteó la necesidad del fortalecimiento de los sistemas de información y vigilancia de la salud materna y se creó dentro de los indicadores de monitoreo y evaluación el registro sistemático de los riesgos o criterios de inclusión materna y la medición de los indicadores del evento [4].

Colombia cuenta con un sistema de información de vigilancia en salud pública estructurado, con una notificación rutinaria a través de un reporte individual de casos y enlace de reportes inmediatos a través de plataformas web, lo cual hace que sea viable la implementación de mecanismos tecnológicos de alerta que enlacen las instituciones de salud, las entidades territoriales de salud y las empresas administradoras de planes de beneficio, optimizando los tiempos de respuesta en busca de respuestas oportunas hacia la obtención de los mejores resultados en la búsqueda de una salud materna y perinatal de alta calidad [4].

Según Popular Science un grupo de científicos de la Universidad de California Los Ángeles (UCLA) desarrolló una técnica que combina un microscopio especial con un algoritmo de inteligencia artificial que funciona para detectar cáncer en muestras de sangre.

Este método combina dos componentes que fueron inventados en la misma universidad: un microscopio que es capaz de tomar fotos en las células de muestras de sangre rápidamente en cuestión de nanosegundos y una técnica de aprendizaje automático 'Deep learning' que identifica células con cáncer que acierta en un 95% de las veces según un reporte de (UCLA) [5].

Este tipo de investigaciones nos demuestra que la inteligencia artificial puede resultar siendo un excelente aliado de la rama de la ciencia de la salud en este caso ya que permite tener un alto grado de aceptabilidad al momento de obtener soluciones eficaces y resulta relevante para nuestro proyecto puesto que el propósito del mismo se basa en desarrollar una red neuronal a través de una de las

técnicas que nos proporciona los algoritmos de inteligencia artificial, en este caso 'Deep Learning' enfocado a la predicción de casos de Morbilidad de Materna Extrema (MME), gracias a esta técnica podemos analizar grandes volúmenes de datos y posteriormente realizar predicciones a través de los resultados obtenidos. Con la elaboración de este proyecto se espera que a futuro la CMRC pueda llegar hacer un referente regional y nacional en atención prenatal y contando con un algoritmo de inteligencia artificial para la prevención de riesgos de MME, permitiendo que otras entidades prestadoras de salud remitan a sus pacientes a la E.S.E CMRC, para realizar procesos confiables a mujeres en estado de embarazo, parto o puerperio prestando un servicio de calidad y contando herramientas tecnológicas para el beneficio de la entidad prestadora de salud y pacientes [5].

1.2. Objetivos

General

Desarrollar una red neuronal profunda para la predicción de riesgo de Morbilidad Materna Extrema (MME) en la etapa del postparto en la E.S.E Clínica de Maternidad Rafael Calvo.

Específicos

- Construir el conjunto de datos con los valores de cada una de las variables de riesgo de Morbilidad Materna Extrema.
- Entrenar la red neuronal usando el 70% del conjunto de datos, permitiendo así definir los parámetros del modelo que mejor representa el comportamiento de los datos.

- Validar el modelo construido usando el 30% del conjunto restante para realizar los ajustes necesarios de los parámetros del modelo en términos del error.

1.3. Alcance del proyecto

Este proyecto se centra en el desarrollo de una red neuronal profunda (Deep Learning) que evalúa mediante algoritmos de Inteligencia Artificial, procedimientos estadísticos y matemáticos con ayuda del lenguaje de programación Python y sus respectivas librerías.

Se utilizarán los datos suministrados por la entidad de salud Clínica de Maternidad Rafael Calvo (CMRC), con fines investigativos de manera confidencial, que datan desde el año 2016 hasta el año 2018 con la población del departamento de Bolívar las cuales hayan sido pacientes remitidas a la CMRC, determinando los factores de riesgo de MME, comprendiendo los datos obtenidos y brindando una predicción de alto, medio, bajo riesgo de MME, contribuyendo así con la reducción de la tasa de Mortalidad Materna Extrema en el departamento de Bolívar.

Con el conjunto de datos (Dataset) obtenido por medio de la Clínica de Maternidad Rafael Calvo (CMRC) serán utilizados como un conjunto de datos obteniendo de ellos variables dependientes e independientes o variables de entrada y salida con los datos bien estructurados y balanceados. Se utilizarán datos reales de la clínica para el entrenamiento y la prueba con el fin de mejorar el aprendizaje y obtener salidas con valores acertados.

Para realizar el entrenamiento de un algoritmo de aprendizaje profundo se necesita de un gran volumen de datos y potencia computacional para poder procesar toda la información requerida por el conjunto de datos. Por lo que se utilizará una máquina virtual de Google llamada Google Colab con alta capacidad de procesamiento.

La predicción que realice este algoritmo de redes neuronales profundas proporcionara una base o soporte para que el personal médico pueda tomar

decisiones, ya que el algoritmo no reemplazara la toma de decisión de un médico. Le servirá para observar si una paciente tiene alto, medio, bajo riesgo de sufrir algún criterio de inclusión catalogándose como MME.

1.4. Marco de referencia

A continuación, se detallan los antecedentes, conceptos y teorías necesarios para comprender el contexto del proyecto.

1.4.1. Estado del arte

En la ciudad de Guerrero-México, al menos tres mujeres embarazadas mueren cada mes, lo que representa una tasa de 51.7%, superando el promedio nacional; esta región ocupa el segundo lugar en términos de muerte materna a nivel nacional. Situación que llevó a PROESAH, asociación civil conformada por un grupo de doctores, profesores y alumnos de la Universidad Autónoma de Guerrero, a desarrollar LUZI, una aplicación que utiliza tecnología de Inteligencia Artificial (IA) de IBM para analizar el riesgo obstétrico [6].

La tecnología de IA de IBM Watson nos está ayudando a monitorear de forma más fácil y rápida los embarazos y a tomar decisiones basadas en evidencia, de acuerdo con las normas y prácticas oficiales de la Secretaría de Salud. LUZI ofrece más de 270 mediciones durante todo el embarazo y ayuda a las mujeres a tener información diaria sobre su estado actual. Según la Organización Mundial de la Salud es posible reducir la mortalidad de las madres con un monitoreo continuo”, dijo la doctora Andrea Sarabia, integrante del grupo de especialistas que desarrollaron LUZI [7].

En otros lugares del mundo se están implementando tecnologías de inteligencia artificial, como en el caso de unos algoritmos capaces de mejorar el diagnóstico de cáncer de mama, lo que hace apuntar a este tipo de investigaciones en redes

neuronales que cada vez más están interesadas en aplicaciones biomédicas, con el objetivo de mejorar el diagnóstico y el tratamiento de enfermedades.

Su iniciativa forma parte del proyecto internacional Digital Mammography DREAM Challenges, considerada como una de las iniciativas más importantes para aplicar la inteligencia artificial en la detección de tipos de tumores. El sistema basado en inteligencia artificial, por el contrario, trata de reducir el número de regiones bajo sospecha y ofrecer información sobre la posible presencia de un cáncer gracias a la utilización de algoritmos predictivos y redes neuronales.

Otro caso famoso fue en el año 2017 el científico nigeriano Charles Onu desarrolló Ubenwa un sistema de aprendizaje automático que puede tomar como entrada el llanto del bebé, analizar los patrones de amplitud y frecuencia en el llanto, para proporcionar un diagnóstico instantáneo de asfixia al nacer. Los resultados de las pruebas del software de diagnóstico demostraron una sensibilidad superior al 86% y una especificidad del 89%.

Es importante agregar que este proyecto se desarrolla en el marco de un proyecto de investigación que busca el análisis de la calidad de los datos y de la precisión de un modelo de predicción para identificar tempranamente riesgos asociados a la ocurrencia de morbilidad materna extrema, este proyecto es liderado por la investigadora Eugenia Arrieta [8]. Quien ha realizado previamente una publicación referente a esta misma temática, pero usando la técnica de regresión logística [9].

1.4.2. Marco teórico

A continuación, se hace una explicación de las definiciones de los términos aplicados en el proyecto, los cuales ayudarán a enmarcar este contexto investigativo, permitiendo ayudar a comprender mucho mejor la temática del proyecto.

Morbilidad Materna Extrema (MME):

La Organización Mundial de la Salud (OMS) define la Morbilidad Materna Extrema (MME) como el caso de una mujer que sobrevive a una complicación de alto riesgo ocurrida durante el embarazo, parto y puerperio. A demás debe cumplir con al menos uno de los criterios de inclusión establecidos [5].

El fallecimiento de una mujer en el momento del parto ha sido durante mucho tiempo el punto de partida para investigar y mejorar los procesos en la salud materna para garantizar un servicio de calidad a las pacientes en los diferentes entes de salud territoriales en Colombia. El conocimiento de las circunstancias que conllevan a la muerte de una paciente ayuda a identificar esos diferentes factores de riesgo causantes de MME relacionados con aspectos clínicos, hereditarios, VIH o diferentes tipos de infecciones que pueden colocar en riesgo de muerte la vida de la paciente categorizándola como Morbilidad Materna Extrema.

Por medio de un protocolo diligenciado por los médicos encargados de atender el proceso de embarazo o el parto, ellos transmiten una ficha técnica para que una persona encargada lo suba a la plataforma de SIVIGILA (Sistema Nacional de Vigilancia en Salud Pública) creado para realizar la recolección en forma sistemática y oportuna de información sobre el comportamiento de los eventos que afecten o puedan afectar la salud de la población colombiana [10].

La tabla de criterios de inclusión materna para la Morbilidad Materna Extrema (MME) en Colombia (ver Tabla 1).

Criterios de inclusión de MME en Colombia	
Diagnóstico clínico	Eclampsia
	Preeclampsia severa
	Sepsis o infección sistémica severa
	Hemorragia obstétrica severa
	Ruptura uterina
Disfunción orgánica	Falla vascular
	Falla renal
	Falla hepática
	Falla metabólica
	Falla cerebral
	Falla respiratoria
	Falla de coagulación
Manejo	Necesidad de transfusión
	Necesidad de UCI
	Necesidad de procedimiento quirúrgico de emergencia

Tabla 1 Criterios de inclusión de MME - Fuente propia

Descripción de los criterios de inclusión MME en Colombia

Cuando una paciente presenta al menos uno de los criterios de inclusión presentado en la tabla 1 es considerado o clasificado como Morbilidad Materna Extrema.

Nombre Variable	Descripción	Tipo de Dato	Valores
Edad	Edad de la paciente	Numérico	0 - 100
Paridad	Indica el número de parto que ha tenido	Numérico	0 - 20
Cesáreas	Es un tipo de intervención quirúrgica el cual se realiza una incisión quirúrgica en el abdomen y el útero de la madre para extraer uno o más bebés.	Binario	0 - 1
Partos Vaginales	Forma tradicional de parto	Numérico	0 - 20

Abortos	Interrupción y finalización prematura del embarazo.	Binario	0 - 1
Molas	Trastorno del embarazo debido a un crecimiento anormal del feto en el útero.	Binario	0 - 1
Ectópicos	El embarazo ectópico es lo que sucede cuando el embarazo avanza fuera del útero, generalmente, en las trompas de Falopio.	Binario	0 - 1
Eclampsia	Aparición de convulsiones o coma durante el embarazo en una mujer.	Binario	0 - 1
Preclamsia	Algunos síntomas de Preeclampsia pueden incluir la hipertensión arterial y proteína en la orina que ocurre después de la semana 20 del embarazo	Binario	0 - 1
Número de Control Prenatal	En cada visita de cuidado prenatal su profesional controla su salud y la de su bebé en crecimiento.	Binario	0 - 1
Sepsis	Ocurre cuando el cuerpo tiene una abrumadora respuesta inmunitaria a una infección bacteriana	Binario	0 - 1
Falla Cardíaca	Durante la gestación ocurren cambios hemodinámicos importantes, que pueden llevar a la descompensación de la paciente con alteración estructural cardíaca previa.	Binario	0 - 1
Falla Renal	La insuficiencia renal materna puede causar, restricción del crecimiento fetal, feto muerto	Binario	0 - 1
Falla Hepática	La frecuencia de estas enfermedades es escasa, pero en ocasiones pueden revestir gravedad tanto para la madre como para el feto.	Binario	0 - 1
Falla Metabólica	Durante la gestación tienen lugar adaptaciones metabólicas maternas para aumentar la sensibilidad a la insulina y mantener un adecuado nivel de glucosa para transferirse al feto.	Binario	0 - 1
Ruptura Uterina	La rotura uterina sucede cuando existe la pérdida de integridad de la pared del útero.	Binario	0 - 1
Hemorragia Obstétrica Severa	Es un sangrado que puede aparecer por los genitales externos o bien, más peligrosamente, una hemorragia intraabdominal.	Binario	0 - 1

Síndrome de Distress	Durante el embarazo es infrecuente, su letalidad es elevada y ha sido reportado como un factor importante en la mortalidad materna.	Binario	0 - 1
Coagulación	Durante el embarazo existe el riesgo de padecer hemorragias uterinas impredecibles y múltiples abortos consecutivos.	Binario	0 - 1
Estuvo en UCI	Si la paciente ingreso o no a la unidad de cuidado intensivos UCI	Binario	0 - 1
Días de estancia en UCI	Días que estuvo en la unidad de cuidados intensivos UCI	Numérico	0 - ...
Transfusión	Si la paciente recibió o no transfusión de sangre por algún motivo alguno	Binario	0 - 1
Diabetes gestacional	Cuando se tiene alta la azúcar en la sangre, generalmente se diagnostica entre las 24 y las 28 semanas de embarazo.	Binario	0 - 1
Anemia	Es una de las complicaciones más frecuentes durante el embarazo, es causada por un descenso del hierro por debajo de los valores normales.	Binario	0 - 1

Tabla 2 Definición de criterios de inclusión de MME - Fuente propia

Redes neuronales profundas (Deep Learning)

El Deep Learning o Aprendizaje profundo es un método del Machine Learning que permite entrenar una Inteligencia Artificial para poder obtener una predicción dado un conjunto de entradas. Esta Inteligencia logrará un nivel de cognición por jerarquías. Se puede utilizar Aprendizaje Supervisado o No Supervisado.

Para lograr obtener predicciones se deben utilizar variables de entradas (Input Layer), se pueden utilizar muchas entradas, la predicción se refleja en la variable de salida (Output Layer). Todas las neuronas (Hidden Layers) están interconectadas unas con otras en las distintas capas (ver Ilustración 1).

- La capa de entrada recibe los datos de entrada y los pasa a la primera capa oculta.
- Las capas ocultas realizarán cálculos matemáticos con nuestras entradas. Uno de los desafíos al crear la Red Neuronal es decidir el número de capas ocultas y la cantidad de neuronas de cada capa.
- La capa de Salida devuelve la predicción realizada.

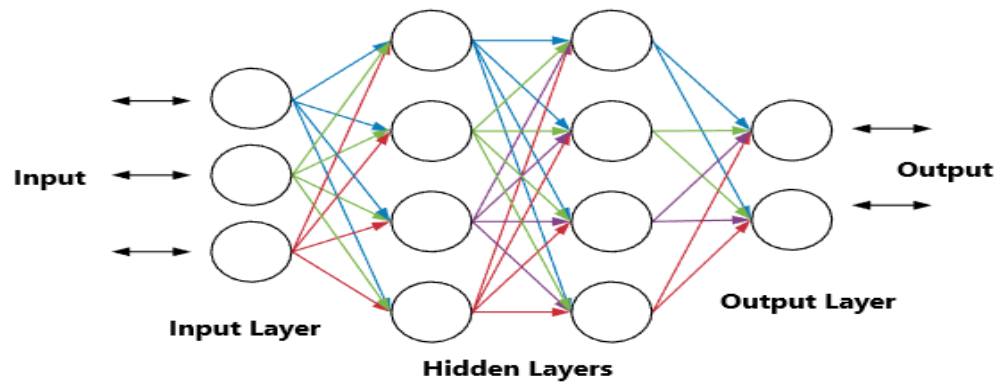


Ilustración 1 Red neuronal organizada en capas. - Fuente propia

Como calcular la predicción

Cada conexión de nuestra red neuronal está asociada a un peso. Este peso dictamina la importancia que tendrá esa relación en la neurona al multiplicarse por el valor de entrada. Los valores iniciales de peso se asignan aleatoriamente y más adelante los pesos se ajustarán solos. (Ver Ilustración 2).

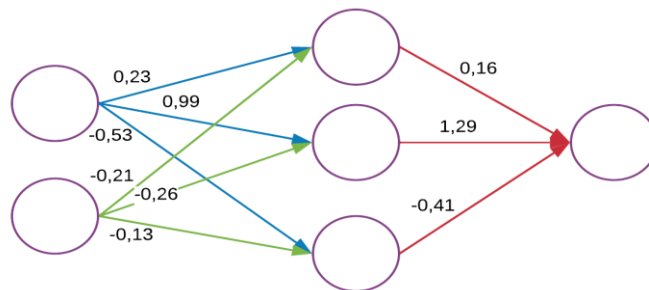


Ilustración 2 Distribución de pesos en una red neuronal - Fuente Propia

Imitando a las neuronas biológicas cada neurona tiene una función de activación. Esta función determinará si la suma de sus valores recibidos (previamente multiplicados por el peso de la conexión) supera un umbral que hace que la neurona se active y dispare un valor hacia la siguiente capa conectada. Cuando todas las capas finalizan de realizar sus cálculos, se llegará a la capa final o de salida con una predicción.

Entrenamiento de una red neuronal:

Entrenar la Inteligencia Artificial puede llegar a ser la parte más difícil del Deep Learning. Se requiere que:

- Gran cantidad de valores en el conjunto de Datos de Entrada
- Gran poder de cálculo computacional

Para entrenar la máquina, se debe alimentar con un conjunto de datos de entrada y comparar el resultado.

Con el conjunto de datos, se comienza un proceso iterativo: donde se usa una función para comparar cuan bueno/malo fue el resultado contra el resultado real. Esta función es llamada Función Coste, idealmente se busca que el coste sea cero, es decir sin error.

A medida que se entrena el modelo irá ajustando los pesos de interconexiones de las neuronas de manera automática hasta obtener buenas predicciones. A ese proceso de ir y venir por las capas de neuronas se le conoce como Back-Propagation.

Redes de propagación hacia atrás (backpropagation)

El nombre de backpropagation resulta de la forma en que el error es propagado hacia atrás a través de la red neuronal, en otras palabras, el error se propaga hacia atrás desde la capa de salida. Esto permite que los pesos sobre las conexiones de las neuronas ubicadas en las capas ocultas cambien durante el entrenamiento.

El cambio de los pesos en las conexiones de las neuronas además de influir sobre la entrada global influye en la activación y por consiguiente en la salida de una neurona. Por lo tanto, es de gran utilidad considerar las variaciones de la función activación al modificarse el valor de los pesos. Esto se llama sensibilidad de la función activación, de acuerdo con el cambio en los pesos.

Cómo reducir la función coste y mejorar las predicciones

Para poder ajustar los pesos de las conexiones entre neuronas haciendo que el coste se aproxime a cero se usa una técnica llamada Gradient Descent. Esta técnica permite encontrar el mínimo de una función.

Funciona cambiando los pesos en pequeños incrementos luego de cada iteración del conjunto de datos. Al calcular la derivada (o gradiente) de la Función Coste en un cierto conjunto de pesos, se logra ver en qué dirección «descender» hacia el mínimo global. (ver Ilustración 3)

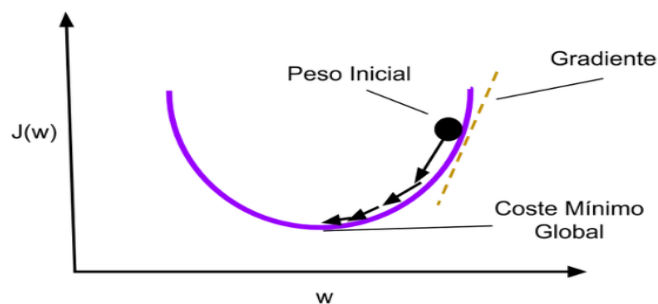


Ilustración 3 Función para reducir la función de coste y mejorar predicciones

Para minimizar la función de coste se necesita iterar por el conjunto de datos cientos de miles de veces (o más), por eso es tan necesario tener gran capacidad de cómputo en el ordenador/nube en la que se entrene la red.

La actualización del valor de los pesos se realizará automáticamente usando el Descenso de Gradiente.

RFE (Recursive Feature Elimination)

Es un clasificador de funciones con eliminación recursiva, su funcionamiento es básicamente una selección hacia atrás de los predictores. Esta técnica comienza construyendo un modelo en todo el conjunto de predictores y calculando un puntaje de importancia para cada predictor. Luego se eliminan los predictores menos importantes, se reconstruye el modelo y se vuelven a calcular los puntajes de importancia.

Penalización

En este modelo fue necesario penalizar la pérdida (loss) del modelo para las muestras de esta clase durante el entrenamiento por medio de la función (*Class_weight*) para lograr un equilibrio con las clases minoritaria de los datos.

Regularización

El aprendizaje consiste en encontrar los coeficientes que minimizan una función de coste. La regularización consiste en añadir una penalización a la función de coste. Esta penalización produce modelos más simples que generalizan mejor. Las regularizaciones más usadas en machine learning son: Lasso (también conocida como L1), Ridge (conocida también como L2) y ElasticNet que combina tanto Lasso como Ridge.

Agregar capas al modelo

Con el método (*.add*) se logra añadir una nueva capa al modelo y pasa a crear una instancia con (*tf.keras.layers.Dense*) donde (*Dense*) implementa la operación del optimizador (*activation=tf.nn.sigmoid*) aplicando la función de activación sigmoide, transformando los valores introducidos a una escala (0,1) en el conjunto de datos.

Función sigmoide:

La función sigmoide transforma los valores introducidos a una escala (0,1), donde los valores altos tienen de manera asintótica a 1 y los valores muy bajos tienden de manera asintótica a 0.

Parámetros para ajustar el modelo de entrenamiento

A partir del modelo (*sequential*) y una vez se tenga definido el mismo, se comienza a configurar el proceso de aprendizaje por medio del método *compile()*, con el que se puede especificar algunas propiedades a través de argumentos del método.

El primero de estos argumentos es la función de (*loss*) que se usará para evaluar el grado de error entre salidas calculadas y las salidas deseadas de los datos de entrenamiento. Se especifica el algoritmo de optimización como Adam, Sigmoide, RMSprop, Adagrad los cuales permiten a la red neuronal calcular los pesos de los parámetros a partir de los datos de entrada y de la función de (*loss*) definida.

Y finalmente se debe indicar la métrica (*accuracy*) que se usara para monitorear el proceso de aprendizaje (*y_test*) de la red neuronal.

Optimizador Adam

Es utilizado para calcular un nuevo valor de cada uno de los parámetros (pesos y sesgos) del modelo en cada iteración de tal forma de que se reduzca el (*loss*). La

optimización de Adam es un método de descenso de gradiente estocástico que se basa en la estimación adaptativa de los momentos de primer y segundo orden.

Normalización de los datos:

A través de la función (MinMaxScaler()) se transforma a escalar cada característica a un rango dado en el conjunto de datos y con la función (fit_transform) se ajustan los datos para luego transformarlos y normalizarlos ya luego con la función (np.array) en forma de matriz.

Esta normalización permitió que las operaciones matriciales que se realizaron en entrenamiento sean mucho más rápidas y eficientes tanto para la multiplicación que se realiza con los pesos que se inicializan y para la actualización de pesos con respecto a nivel de error que se debe disminuir.

1.4.3. Marco conceptual

Redes neuronales:

Es un modelo que, inspirado en el cerebro, está compuesto de capas (al menos una de las cuales está oculta) que consisten en unidades conectadas simples o neuronas seguidas de no linealidades.

Arquitectura de redes neuronales:

Se nombra arquitectura a la topología, estructura o patrón de conexión de una red neuronal. En una red neuronal artificial todos los nodos se conectan por medio de sinapsis, estando el comportamiento de la red determinado por la estructura de conexiones sinápticas.

Las conexiones sinápticas son direccionales, lo que quiere decir, que la información solamente puede orientarse en un único sentido (desde la neurona presináptica a

la postsináptica). En un contexto generalizado las neuronas se suelen agrupar en unidades estructurales que se denominan capas. Es el conjunto de una o más capas constituyen como tal la red neuronal [11].

Se distinguen tres tipos de capas:

- Capa de entrada
- Capa de salida
- Capas ocultas

Funciones de base y activación:

El valor de red, expresado por la función de base, $u(w, x)$, se transforma mediante una función de activación no lineal. Las funciones de activación más comunes son la función sigmoideal y Gaussiana:

- *Función Sigmoideal:* $f(u_i) = \frac{1}{1 + \exp\left\{\frac{u_i}{\sigma^2}\right\}}$
- *Función Gaussiana:* $f(u_i) = c \exp\left\{-\frac{u^2}{\sigma^2}\right\}$

Función de activación:

La función de activación se encarga de devolver una salida a partir de un valor de entrada, normalmente el conjunto de valores de salida en un rango determinado como (0,1) o (-1,1) [12].

Función Relu (Rectified Lineal Unit)

La función ReLU transforma los valores introducidos anulando los valores negativos y dejando los positivos tal y como entran.

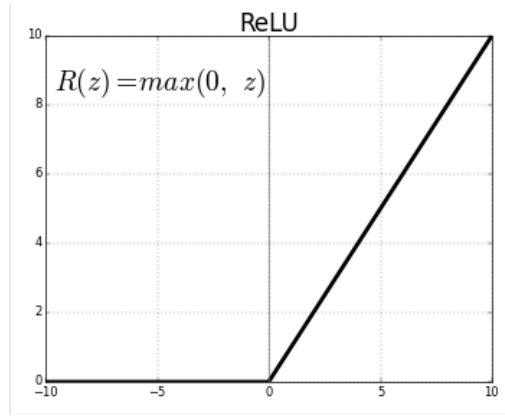


Tabla 3 Función de activación Relu

Softmax:

La función Softmax transforma las salidas a una representación en forma de probabilidades, de tal manera que el sumatorio de todas las probabilidades de las salidas de 1.

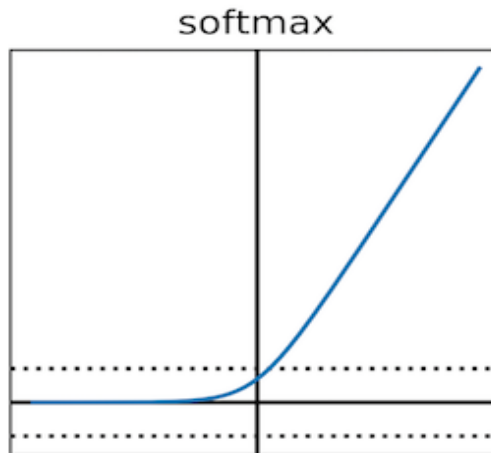


Tabla 4 Función de activación Softmax

AUC (Área bajo la curva):

Es la métrica de evaluación que considera todos los umbrales de clasificación posibles. El área bajo la curva ROC es la probabilidad de que un clasificador tenga más seguridad de que un ejemplo positivo elegido al azar sea realmente positivo con respecto a que un ejemplo negativo elegido al azar sea positivo.

Conjunto de datos (data set):

Es el total del conjunto de datos sobre los cuales se quiere desarrollar un algoritmo de aprendizaje automático con el fin de obtener un modelo que lo represente lo mejor posible. Posee variables dependientes e independiente.

Variables dependientes (Labels):

Es la columna del dataset que responde a una correlación de variables independientes y que debe ser predicha por el futuro modelo en este caso la columna MME.

Variables independientes (Features):

Son aquellas columnas que hacen parte del dataset que serán usadas por el algoritmo para generar un modelo que prediga lo mejor posible las variables independientes.

Conjunto de datos de entrenamientos (Training set):

Es el subconjunto del dataset que será utilizado para entrenar el modelo que se pretende generar.

Conjunto de datos de prueba (Test set):

Es el subconjunto del dataset que se le pasara al modelo una vez haya sido entrenado para comprobar, mediante el uso de diferentes métricas, su calidad.

1.4.4. Consideraciones éticas y marco legal

Consideraciones éticas:

Al relacionar la Inteligencia Artificial con los procesos organizacionales, es bien importante dotar a esta nueva tecnología de valores y principios. Y, dentro de las organizaciones, son los desarrolladores de esta tecnología las personas que realmente han de trabajar y desarrollar conscientes de las implicaciones morales y éticas que conlleva un buen trabajo.

Las grandes organizaciones, empresas y gobiernos se están centrando en los problemas que pueden surgir con el tema de la ética de la inteligencia artificial para trazar consideraciones, prácticas y marcos comunes de cara al futuro. Es importante conseguir un acuerdo donde se pueda conceptualizar, pero, sobre todo, regular las prácticas derivadas y asegurar que la IA este centrada en el bien para el ser humano.

Los estándares morales dirigidos a humanos, para los creadores de tecnología los principios son los siguientes:

- Debe respetar los derechos fundamentales y la regulación aplicable.
- Debe ser técnicamente robusta y fiable.
- Debe funcionar con transparencia.
- No debe restringir la libertad humana.

Caso de implementación:

Al realizar la implementación de la Inteligencia Artificial en las distintas áreas se deben tener en cuenta las consideraciones éticas, en este caso, las redes neuronales benefician los procesos que diariamente realizan los seres humanos y que se aplican en los diferentes entornos laborales como la economía, los sistemas, la industria, la medicina como en el caso de este proyecto que consta de un modelo de red neuronal profunda (Deep Learning) para predecir riesgos de Morbilidad Materna Extrema, realizado en la Universidad del Sinú.

- **Marco Legal**

En Colombia la ley 100 de 1993:

Establece la prioridad a la infancia se da a través de la atención de las madres gestantes, la población infantil menor de un año y las mujeres de familia.

El decreto 1011 de 2006:

Por el cual se establece el Sistema Obligatorio de Garantía de Calidad de la Atención de Salud del Sistema General de Seguridad Social en Salud.

La circular Externa 030 de 2006:

La Superintendencia Nacional de Salud: Da instrucciones en materia de indicadores de calidad para evaluar la oportunidad, accesibilidad, continuidad, pertinencia y seguridad en la prestación de los servicios de salud de las instituciones prestadoras de servicios de salud [13].

La ley 1122 de 2007:

Por la cual se hacen algunas modificaciones en el Sistema General de Seguridad Social en Salud y se dictan otras disposiciones.

La ley 1341 de 30 Julio de 2009:

El Congreso de la Republica de Colombia redacta la ley por la cual se definen los principios y conceptos sobre la sociedad de la información y la organización de las Tecnologías de la Información y las Comunicaciones (TIC), se crea la Agencia Nacional de Espectro y se dictan otras sentencias [11].

Hábeas Data (Ley 1581 de 2012):

La Superintendencia de Industria y Comercio tiene plenas facultades para ejercer el control y vigilancia sobre los responsables y encargados de las bases de datos personales.

Esta es una norma estatutaria que desarrolla el derecho fundamental de Hábeas Data consagrado desde 1991 en la Constitución Política, e instrumenta los mecanismos que les permiten a todos los colombianos hacer efectivo sus derechos de acceso, actualización, rectificación y supresión de datos personales ante cualquier entidad que administre base de datos [10].

Agencia Nacional del Espectro (ANE):

Es la entidad que se asume el compromiso de planear con estrategia el uso del espectro radioeléctrico, así como la observación y control en todo el territorio nacional colombiano [11].

Licencia de maternidad:

Corresponde a uno de los derechos más importantes para las trabajadoras en materia de seguridad social. Consiste en el tiempo de descanso remunerado, prenatal como postparto. En Colombia, el permiso remunerado o de licencia por maternidad hasta inicios del 2017 era de 14 semanas. Por la ley 1822 del 4 de enero de 2017 se extendió a 18 semanas [11].

Ministerio de Salud y protección Social. Resolución 603280 del 2 de agosto de 2018:

Por medio de la cual se adoptan los lineamientos técnicos y operativos de la Ruta Integral de Atención para la Promoción y Mantenimiento de la Salud y la Ruta Integral de Atención en Salud para la Población Materno Perinatal [11].

Las TIC deberán colaborar, dentro del marco de sus obligaciones, para priorizar el acceso y uso a las Tecnologías de la Información y las Comunicaciones en la producción de bienes y servicios, en condiciones no discriminatorias en la conectividad, la educación, los contenidos y la competitividad [11].

1.5. Metodología

1.5.1. Línea de investigación

La universidad del Sinú seccional Cartagena consta de diversos grupos de investigación, con el fin de aportar en avances tecnológicos logrando así la

optimización de procesos y encaminados a la generación y aplicación de conocimiento relevante. La escuela de ingeniería de sistemas cuenta con el grupo de investigación DEARTICA, dentro de las líneas de investigación se encuentra inteligencia artificial, la cual es la línea en la que va encaminada este proyecto.

1.5.2. Tipo de investigación

La investigación de este proyecto tiene un enfoque metodológico mixto, en el que se aborda desde una concepción inicial cualitativa y se complementa con una concepción cuantitativa. Este es un proyecto de desarrollo tecnológico que incluye el uso de técnicas de inteligencia artificial usando lenguaje de programación Python. El nivel de innovación del proyecto es medio por su aplicabilidad al sector salud y aporte en beneficio de la salud de la población del departamento de Bolívar.

Los datos que se usarán serán proporcionados por la Clínica de Maternidad Rafael Calvo y serán datos anónimos para garantizar la confidencialidad e integridad de los pacientes.

Es importante aclarar que este proyecto es una extensión del proyecto de investigación macro “Predicción Temprana de Morbilidad Materna Extrema usando Aprendizaje Automático” a cargo de la Coordinadora de investigación de la escuela de ingeniería de sistemas.

1.5.3. Descripción de la metodología

En la elaboración de este proyecto se utilizará la metodología Team Data Science Process (TDSP), que traducido al español significa Proceso de Ciencia de Datos del Equipo, la cual consiste en brindar un resultado de manera eficiente mediante el análisis predictivo, abarca las mejores prácticas que nos brinda Microsoft, mejorando el aprendizaje de forma colectiva permitiendo así simplificar la creación y la innovación en ideas en la ciencia de datos [14].

Ciclo de vida de la ciencia de datos

Para formar parte de las aplicaciones inteligentes que implementan modelos de aprendizaje automático o inteligencia artificial para el análisis predictivo. Otras clases de proyectos como los exploratorios de ciencias de datos y análisis ad hoc donde también se pueden ver beneficiados [12].

El ciclo de vida describe las etapas principales que los proyectos suelen ejecutar, a menudo de forma iterativa:

1. Conocimiento del negocio
2. Adquisición y comprensión de datos
3. Modelado
4. Despliegue
5. Aceptación del cliente

(El alcance de este proyecto llegará hasta el 3 punto del ciclo de vida)

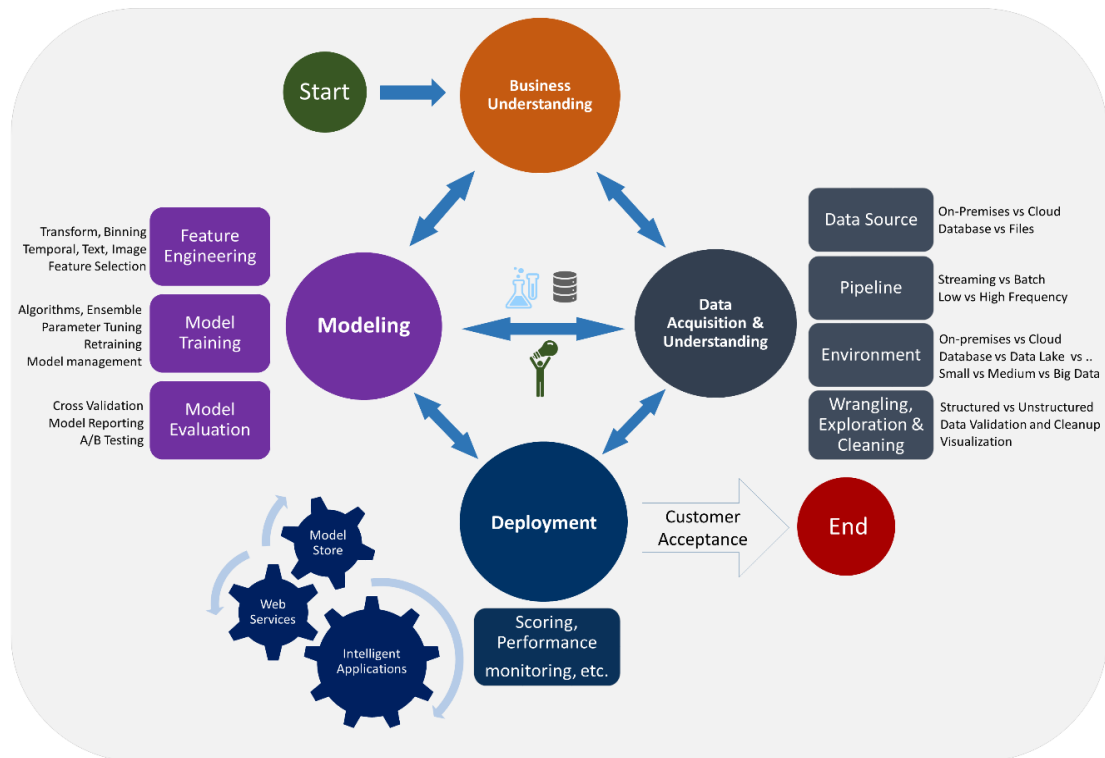


Ilustración 4 Ciclo de vida metodología TDSP de Microsoft [15]

Tomando como referencia esta metodología se presenta la siguiente matriz metodológica, donde se estructuran una serie de fases y se define una clara ruta de investigación, con ello se busca planificar el desarrollo de las actividades que posibiliten la consecución de los objetivos específicos, y desde allí al logro del objetivo general.

Fases	Objetivos Específicos	Actividades
Conocimiento del negocio	Construir el conjunto de datos con los valores de cada una de las variables de riesgo de Morbilidad Materna Extrema. (Regresión lineal múltiple)	<ul style="list-style-type: none"> • Entrevistas con ginecólogos • Análisis del Protocolo de Morbilidad Materna Extrema • Análisis de la consulta prenatal • Investigación teórica sobre los criterios de inclusión en Colombia
Modelado	Entrenar la red neuronal usando el 70% del conjunto de datos, permitiendo así definir los parámetros del modelo que mejor representa el comportamiento de los datos.	<ul style="list-style-type: none"> • Aplicación de técnicas de filtrado de variable • Normalización de los datos • División de los datos • Balanceo de los datos • Construcción del modelo • Entrenamiento del modelo
	Validar el modelo construido usando el 30% del conjunto restante para realizar los ajustes necesarios de los parámetros del modelo en términos del error.	<ul style="list-style-type: none"> • Evaluación del modelo • Aplicación de métricas de evaluación

Tabla 5 Cumplimiento de los objetivos con base a la metodología

2. CONOCIMIENTO DEL NEGOCIO

En este capítulo se expone el conocimiento sobre los procesos ocurridos en la CMRC en donde se realizan y monitorean a las maternas. Gracias al valioso aporte brindado por el Doctor Francisco Edna se conocerá como es la atención de una paciente que comienza en su etapa uno del embarazo, el conocimiento de citas prenatales y al proceso que debe tener una paciente dictaminada con algún criterio de inclusión materna, catalogándose como MME en la CMRC.

Acorde a las conversaciones sostenidas con el Doctor Francisco Edna, medico Ginecólogo de la CMRC, entrevemos que el sistema de salud colombiano se rige por el Modelo Integral de Atención en Salud (MIAS), el cual consiste en el conjunto de acciones coordinadas, complementarias y efectivas dirigidas a las personas, familias y comunidades, basado en lineamiento técnico perinatal que comienza a partir de la semana uno.

La población materna perinatal se ve regulada por la resolución 3280 que establece las directrices para su operación y adopta los lineamientos técnicos y operativos de la Ruta Integral de Atención para la Promoción y Mantenimiento de la Salud y la Ruta Integral de Atención en Salud.

2.1. Consulta preconcepcional

Este tipo de consulta se basa en la paciente que aún no ha quedado en estado de embarazo, o que desea quedar embarazada, con este fin se busca realizar un conjunto de exámenes para identificar posibles factores de riesgo y mejorar las condiciones de la paciente para que tenga un embarazo lo más sano posible y así disminuir la probabilidad de sufrir algún criterio de inclusión materna.

Entre los exámenes que se practican a la mujer están los hemogramas, VIH, hepatitis, toxoplasma, retrovirus, varicela, los anteriores análisis solo aplica para antes del embarazo, si este dictamen da como resultado negativo, la mujer está queda en condiciones para quedar en estado de embarazo, adicionalmente se

recomienda 3 meses antes tomar ácido fólico de 1mg para reducir el riesgo de enfermedades, si los resultados arrojan positivo para VIH se realiza un tratamiento y cuando este se encuentre en un estado óptimo, la paciente podrá decidir si desea quedar en embarazo o no.

Asimismo, aplica para mujeres que tengan una vida sexual activa y se espera que estén acudiendo a algún programa de planificación en el que cuente con un esquema de vacunación de hepatitis, tétano, varicela;

Los principales objetivos de las consultas preconceptionales:

- Evaluar los riesgos.
- Realizar acciones preventivas
- Recetar la suplementación farmacológica adecuada.

2.2. Proceso de atención

De acuerdo con las apreciaciones emitidas por el Dr. Francisco Edna, Todo comienza con la consulta preconceptional donde se pretende mejorar las condiciones maternas para una futura gestación. Por medio de la historia clínica, exámenes como el hemograma, VIH, hepatitis y la identificación de las condiciones psicosociales de la paciente. Se recomienda realizar a toda mujer que tenga una vida sexual activa.

Si la paciente es ingresada a la clínica con algún tipo de patología se debe seguir un tratamiento de acuerdo con las anomalías que presente durante el estado de gestación. Por ejemplo: Una paciente que es remitida por hipertensión, ingresa como notificación para el protocolo de Morbilidad Materna Extrema código INS: 549.

2.2.1. Protocolo de Morbilidad Materna Extrema código INS: 549.


		SISTEMA NACIONAL DE VIGILANCIA EN SALUD PÚBLICA Subsistema de información SIVIGILA Ficha de notificación individual – Datos complementarios Morbilidad materna extrema. Cod INS 549				
La ficha de notificación es para fines de vigilancia en salud pública y todas las entidades que participen en el proceso deben garantizar la confidencialidad de la información LEY 1273/09 y 1266/09						
RELACIÓN CON DATOS BÁSICOS			FOR-R02.0000-075 V:00 2019/02/01			
A. Nombres y apellidos del paciente		B. Tipo de ID*	C. Número de identificación			
<small>*RC : REGISTRO CIVIL I : TARJETA IDENTIDAD CC : CÉDULA CIUDADANÍA CE : CÉDULA EXTRANJERÍA PA : PASAPORTE MS : MENOR SIN ID AS : ADULTO SIN ID PE : PERMISO ESPECIAL DE PERMANENCIA</small>						
5. SISTEMA DE REFERENCIA						
5.1 ¿La paciente ingresa remitida de otra institución? <input type="radio"/> 1. Sí <input type="radio"/> 2. No		5.2 Institución de referencia 1	5.3 Institución de referencia 2	5.4 Tiempo del trámite de remisión <input type="text"/> <input type="text"/> Horas		
6. CARACTERÍSTICAS MATERNAS						
6.1 Número de gestaciones <input type="text"/> <input type="text"/>	6.2 Partos vaginales <input type="text"/> <input type="text"/>	6.3 Cesáreas <input type="text"/> <input type="text"/>	6.4 Abortos <input type="text"/> <input type="text"/>	6.5 Molas <input type="text"/> <input type="text"/>	6.6 Ectópicos <input type="text"/> <input type="text"/>	6.7 Muertos <input type="text"/> <input type="text"/>
6.8 Vivos <input type="text"/> <input type="text"/>	6.9 Fecha de terminación de la última gestación (dd/mm/aaaa) <input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>		Incluye el embarazo actual o el que terminó en los 41 días anteriores 6.11. Número de controles prenatales <input type="text"/> 6.12 Semanas al Inicio CPN <input type="text"/>			
6.13 Terminación de la gestación <input type="radio"/> 1. Aborto <input type="radio"/> 2. Parto <input type="radio"/> 3. Parto Instrumentado <input type="radio"/> 4. Cesárea <input type="radio"/> 5. Continúa embarazada			6.15 Momento de ocurrencia con relación a terminación de gestación <input type="radio"/> 1. Antes <input type="radio"/> 2. Durante <input type="radio"/> 3. Después			

Ilustración 5 Protocolo de Morbilidad Materna Extrema Cod INS 549 – tomado de <https://www.ins.gov.co/Paginas/Inicio.aspx>

3. NOTIFICACIÓN			
3.1 Fuente <input type="radio"/> 1. Notificación rutinaria <input type="radio"/> 2. Búsqueda activa Inst. <input type="radio"/> 3. Vigilancia Intensificada <input type="radio"/> 4. Búsqueda activa com. <input type="radio"/> 5. Investigaciones		3.2 Departamento y municipio de residencia del paciente Departamento <input type="text"/> <input type="text"/> Municipio <input type="text"/> <input type="text"/> <input type="text"/>	
3.3 Dirección de residencia			
3.4 Fecha de consulta (dd/mm/aaaa) <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	3.5 Fecha de inicio de síntomas (dd/mm/aaaa) <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	3.6 Clasificación inicial de caso <input type="radio"/> 1. Sospechoso <input type="radio"/> 2. Probable <input type="radio"/> 3. Conf. por laboratorio <input type="radio"/> 4. Conf. Clínica <input type="radio"/> 5. Conf. nexa epidemiológico	3.7 Hospitalizado <input type="radio"/> Sí <input type="radio"/> No
3.8 Fecha de hospitalización (dd/mm/aaaa) <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	3.9 Condición final <input type="radio"/> 1. Vivo <input type="radio"/> 2. Muerto <input type="radio"/> 0. No sabe, no responde	3.10 Fecha de defunción (dd/mm/aaaa) <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	3.11 Número certificado de defunción <input type="text"/> <input type="text"/>
3.12 Causa básica de muerte <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>		3.13 Nombre del profesional que diligenció la ficha	3.14 Teléfono
4. ESPACIO EXCLUSIVO PARA USO DE LOS ENTES TERRITORIALES			
4.1 Seguimiento y clasificación final del caso <input type="radio"/> 0. No aplica <input type="radio"/> 3. Conf. por laboratorio <input type="radio"/> 4. Conf. Clínica <input type="radio"/> 5. Conf. nexa epidemiológico <input type="radio"/> 6. Descartado <input type="radio"/> 7. Otra actualización <input type="radio"/> D. Descartado por error de digitación		4.2 Fecha de ajuste (dd/mm/aaaa) <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	

Correos: siviglia@ins.gov.co

Ilustración 6 Protocolo de Morbilidad Materna Extrema Cod INS 549 – tomado de <https://www.ins.gov.co/Paginas/Inicio.aspx>

7. CRITERIOS DE INCLUSIÓN					
7.1. Enfermedad específica					
7.1.1 Eclampsia	<input type="radio"/> 1. Si	<input type="radio"/> 2. No	7.1.4 Hemorragia obstetrica severa	<input type="radio"/> 1. Si	<input type="radio"/> 2. No
7.1.2 Preeclamsia severa	<input type="radio"/> 1. Si	<input type="radio"/> 2. No	7.1.5 Ruptura uterina	<input type="radio"/> 1. Si	<input type="radio"/> 2. No
7.1.3 Sepsis o infección sistémica severa	<input type="radio"/> 1. Si	<input type="radio"/> 2. No			
7.2. Falla orgánica					
7.2.1 Cardíaca	<input type="radio"/> 1. Si	<input type="radio"/> 2. No	7.2.5 Metabólica	<input type="radio"/> 1. Si	<input type="radio"/> 2. No
7.2.2 Vascular	<input type="radio"/> 1. Si	<input type="radio"/> 2. No	7.2.6 Cerebral	<input type="radio"/> 1. Si	<input type="radio"/> 2. No
7.2.3 Renal	<input type="radio"/> 1. Si	<input type="radio"/> 2. No	7.2.7 Respiratoria	<input type="radio"/> 1. Si	<input type="radio"/> 2. No
7.2.4 Hepática	<input type="radio"/> 1. Si	<input type="radio"/> 2. No	7.2.8 Coagulación	<input type="radio"/> 1. Si	<input type="radio"/> 2. No
7.3. Manejo específico					
7.3.1. Ingreso a UCI	<input type="radio"/> 1. Si	<input type="radio"/> 2. No			
7.3.2. Cirugía adicional	<input type="radio"/> 1. Si	<input type="radio"/> 2. No			
7.3.3. Transfusión	<input type="radio"/> 1. Si	<input type="radio"/> 2. No			
TOTAL DE CRITERIOS					
<input style="width: 40px; height: 20px;" type="text"/> <input style="width: 40px; height: 20px;" type="text"/>					
8. DATOS RELACIONADOS CON EL MANEJO					
8.1 Días de estancia hospitalaria	8.3. Si en el numeral 7.3.3 marcó SI indique cuantas unidades transfundidas		8.4. Si en el numeral 7.3.2 marcó SI indique que cirugía		
<input style="width: 40px; height: 20px;" type="text"/>	<input style="width: 100px; height: 20px;" type="text"/>		<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>CIRUGIA 1</p> <input type="radio"/> 1. Histerectomía <input type="radio"/> 2. Laparatomía <input type="radio"/> 3. Legrado <input type="radio"/> 4. Otra </div> <div style="width: 45%;"> <p>CIRUGIA 2</p> <input type="radio"/> 1. Histerectomía <input type="radio"/> 2. Laparatomía <input type="radio"/> 3. Legrado <input type="radio"/> 4. Otra </div> </div>		
8.2. Días de estancia en cuidados intensivos	Unidades transfundidas		Cuál? _____		
<input style="width: 40px; height: 20px;" type="text"/>			Cuál? _____		

Ilustración 7 Protocolo de Morbilidad Materna Extrema Cod INS 549 - 3 - tomado de <https://www.ins.gov.co/Paginas/Inicio.aspx>

El Sistema Nacional de Vigilancia en Salud Pública -SIVIGILA, ha sido creado para realizar la provisión en forma sistemática y oportuna, de información sobre la dinámica de los eventos que afecten o puedan afectar la salud de la población colombiana, con el fin de:

- Orientar las políticas y la planificación en salud pública.
- Tomar las decisiones para la prevención y control de enfermedades y factores de riesgo en salud.
- Optimizar el seguimiento y evaluación de las intervenciones.
- Racionalizar y optimizar los recursos disponibles y lograr la efectividad de las acciones en esta materia, propendiendo por la protección de la salud individual y colectiva.

Una vez la paciente se encuentre en la institución prestadora de salud tiene 2 opciones para acceder al parto vaginal o parto por cesárea, por ambas vías se pueden presentar complicaciones al momento de realizar el procedimiento ya que pueden surgir crisis hipertensivas, hemorragias o cualquier otra patología que de no ser atendida oportunamente prácticamente puede perder la vida el feto o la paciente durante el estado de gestación, parto o puerperio pero logran sobrevivir, se le conoce como Morbilidad Materna Extrema (MME) .

2.3. Consulta prenatal

El control prenatal se define como todas las acciones y procedimientos, sistemáticos o periódicos, destinados a la prevención, diagnóstico y tratamiento de los factores que pueden condicionar la Morbilidad y Mortalidad materna y perinatal.

En una consulta se establecen los objetivos del control prenatal, como lo son:

- Evaluar el estado de salud de la gestante
- Identificar factores de riesgo biopsico-sociales presentes
- Identificar o evaluar enfermedades asociadas y propias de la gestación
- Educar y apoyar a la mujer gestante, así como a su pareja o familiares

Para cumplir estos objetivos el personal encargado debe realizar una serie de actividades como:

- Diligenciar Historia Clínica Materna y llenar el Carné Prenatal
- Solicitar y evaluar los exámenes paraclínicos básicos según edad gestacional
- Evaluar necesidad de otros exámenes diagnósticos
- Formular y suministrar micronutrientes
- Formular y administrar toxoide tetánico y otras vacunas de ser necesario

2.4. Criterios de inclusión

Los criterios de inclusión para tener en cuenta de acuerdo con el protocolo de MME son la eclampsia, preeclampsia, sepsis, hemorragias obstétricas, disfunciones orgánicas como falla renal, falla cardíaca, falla hepática, falla metabólica, cerebral, respiratoria, ruptura uterina, coagulación, mala atención a las pacientes, malos procedimientos en UCI y malos procedimientos quirúrgicos, entre otros criterios mencionados en el Protocolo de Morbilidad Materna Extrema.

Tabla de Criterios de Inclusión Materna para Morbilidad Materna Extrema en Colombia (ver Tabla 6).

Criterios de inclusión de MME en Colombia	
Diagnóstico clínico	Eclampsia
	Preeclampsia severa
	Sepsis o infección sistémica severa
	Hemorragia obstétrica severa
	Ruptura uterina
Disfunción orgánica	Falla vascular
	Falla renal
	Falla hepática
	Falla metabólica
	Falla cerebral
	Falla respiratoria
	Falla de coagulación
Manejo	Necesidad de transfusión
	Necesidad de UCI
	Necesidad de procedimiento quirúrgico de emergencia

Tabla 6 Criterios de inclusión de MME - Fuente propia

3. ADQUISICIÓN Y COMPRENSIÓN DE LOS DATOS

En este capítulo, se detallan las herramientas y métodos que se utilizaron para la obtención de los datos requeridos para el entrenamiento posterior de un modelo de predicción. Iniciando con la verificación de la calidad de los datos y la obtención de los mismo.

3.1. Calidad de los datos

La obtención de los datos se realizó con la aprobación de la entidad de salud Clínica Maternidad Rafael Calvo, a fin de generar el set de datos para el entrenamiento de la red neuronal. Así mismo se debe tener en cuenta la variedad de tipos de datos encontrados y los que soporta un modelo de predicción. En este proceso se encontraron tipos numéricos, texto, decimales y binarios, siendo esta variedad un problema para el entrenamiento, porque el modelo no soporta tipos de datos texto, por lo cual se realizó un proceso de transformación de los datos de ciertas variables categóricas en numéricas.

Para el manejo de estos datos se contó con un motor de bases de datos SQL versión 2008, con una intercalación (*Modern_spanish*) ya que esto permite la compatibilidad del motor para ejecutar la base de datos, adicionalmente con un manejador de base de datos de cualquier versión, para este caso se usó versión 2012.

3.1.1. Obtención de los datos

Inicialmente se identificó las variables de tipo texto, y se usó Microsoft Excel para hacer la transformación de los datos tipo texto a numéricas, mediante la función: `=CONCATENAR ("WHEN nombre_del_campo="";campo;" THEN "; columna_id)` Como se aprecia en la Ilustración 8.

A	B
1	MUNICIPIO
2	1 Arenal
3	2 Soledad
4	3 Chigorodó
5	4 Almeida
6	5 Magangué
7	6 Turbaná
8	7 Santa ana
9	8 Córdoba
10	9 Barrancabermeja


```

WHEN municipio= 'Arenal' THEN 1
WHEN municipio= 'Soledad' THEN 2
WHEN municipio= 'Chigorodó' THEN 3
WHEN municipio= 'Almeida' THEN 4
WHEN municipio= 'Magangué' THEN 5
WHEN municipio= 'Turbaná' THEN 6
WHEN municipio= 'Santa ana' THEN 7
WHEN municipio= 'Córdoba' THEN 8
WHEN municipio= 'Barrancabermeja' THEN 9

```

Ilustración 8 Código SQL – Fuente propia

Luego de haber realizado este proceso de identificar cada uno de los valores de los campos procedemos a adherir el resultado de esta función en la base de datos y ejecutar la sentencia para que los valores de tipo texto queden identificado con un número y trabajar de manera adecuada en la red neuronal. (Ver ilustración 9)

```

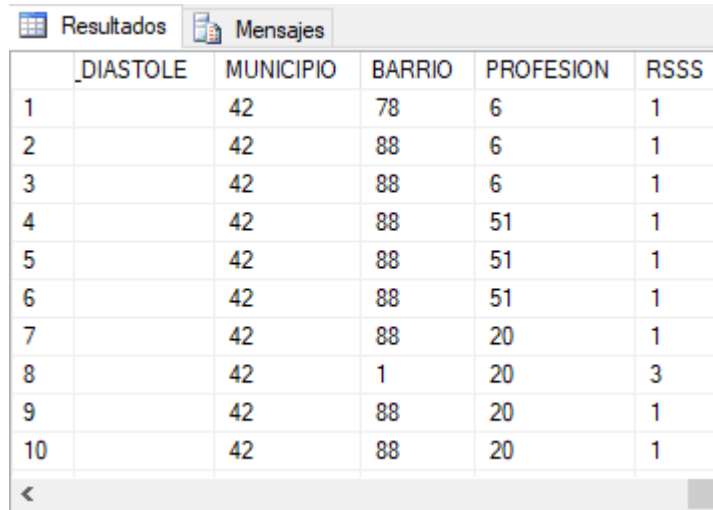
USE [PROYECTO]
GO

SELECT ID,EDAD,ESTRATO, frecuencia_cardiaca AS FRECUENCIA_CARDIACA, frecuencia_respiratoria AS FRECUENCIA_RESPIR
presion AS PRESION, temperatura AS TEMPERATURA, talla AS TALLA, peso AS PESO, imc AS IMC, gestacion AS GESTACION
cesarea AS CESAREA, aborto AS ABORTO, vivos AS VIVOS, PresionSistole AS PRESION_SISTOLE, PresionDiastole AS PRES
CASE WHEN municipio= 'Arenal' THEN 1
WHEN municipio= 'Almeida' THEN 2
WHEN municipio= 'Corozal' THEN 3
WHEN municipio= 'Magangué' THEN 4
WHEN municipio= 'Turbaná' THEN 5
WHEN municipio= 'Santa ana' THEN 6
WHEN municipio= 'Córdoba' THEN 7
WHEN municipio= 'Margarita' THEN 8
WHEN municipio= 'San bernardo del viento' THEN 9
WHEN municipio= 'Repelón' THEN 10

```

Ilustración 9 Código SQL – Fuente propia

El resultado de esta ejecución se puede ver en la ilustración 10, en la cual podemos apreciar que el campo municipio, barrio, profesión y rsss que son variables de tipo texto pasaron a ser de tipo numérico al igual que las demás columnas de la tabla.



	_DIASTOLE	MUNICIPIO	BARRIO	PROFESION	RSSS
1		42	78	6	1
2		42	88	6	1
3		42	88	6	1
4		42	88	51	1
5		42	88	51	1
6		42	88	51	1
7		42	88	20	1
8		42	1	20	3
9		42	88	20	1
10		42	88	20	1

Ilustración 10 Código SQL – Fuente propia

Este proceso es necesario realizarlo con todas las tablas en la que los campos contengan variables de tipo texto, ya que como se indicó en un principio la red neuronal a desarrollar solo trabaja con variables de tipo numérico, así mismo crear una nueva tabla de cada una de estas consultas para posteriormente procesar los datos y darle el manejo adecuado.

Seguidamente se procede a organizar de manera cronológica las nuevas tablas que se han creado a través de las consultas ejecutadas anteriormente mediante la sentencia “*JOIN*” que nos da como resultado lo que se aprecia en la ilustración 11.


```

SELECT e.ID,e.EDAD,e.ESTRATO,e.CODIGO_DIAGNOSTICO_PRINCIPAL,e.CODIGO_DIAGNOSTICO_PRINCIPAL_SALIDA,
e.CODIGO_DIAGNOSTICO_RELACIONADO1,e.CODIGO_DIAGNOSTICO_RELACIONADO1_SALIDA,e.MUNICIPIO,e.BARRIO,e.PROFESION,
e.RSSS,e.ESTADO_CIVIL,e.APLICA_CUIDADO_ENF,e.ESTADO_SALIDA,e.MOTIVO_SALIDA, e.SERVICIO_TEXTO,c.FRECUENCIA_CARDIA
c.FRECUENCIA_RESPIRATORIA,c.PRESION,c.TEMPERATURA,c.TALLA,c.PESO,c.IMC,c.GESTACION,c.PARTO,c.CESAREA,c.ABORTO,
c.VIVOS,c.PRESION_SISTOLE,c.PRESION_DIASTOLE,c.TIPO_DIAGNOSTICO,c.CAUSA_EXTERNA,c.A_OTROS,c.A_FARMACOLOGICOS,c.A
c.A_HOSPITALARIOS,c.A_TRAUMATICOS,c.A_PATOLOGICOS,c.A QUIRURGICOS,c.A_TOXICO,c.A_TRANFUSIONALES,c.A_ALERGIA_MEDI
d.DX_INGRESO1,d.DX_INGRESO2,d.DX_INGRESO3,d.DX_INGRESO4,d.DX_EGRESO1,d.DX_EGRESO2,d.DX_EGRESO3,d.DX_EGRESO4,
p.OBSTETRICOS_GESTACIONES,p.OBSTETRICOS_PARTOS,p.OBSTETRICOS_ABORTOS,p.OBSTETRICOS_CESAREAS,p.MESES_SEMANAS,
p.GRUPO_RH_MADRE,p.CP,p.VIVO_MUERTO,p.SEXO,u.FRECUENCIA_CARDIACA,u.FRECUENCIA_RESPRATORIA,u.PRESION,
u.TEMPERATURA,u.IMC,u.A_FARMACOLOGICOS,u.A_GINECOOBSTETRICO,u.A_HOSPITALARIOS,u.A_TRAUMATICOS,u.A_PATOLOGICOS,
u.A_TOXICO, u.A_TRANFUSIONALES,u.CONTRAREFERENCIA,u.EMBARAZADA,u.CAFALES,u.VOMITOS,u.ADEMA,u.SANGRADO_VAGINAL,
u.EPIGASTRALGIA,u.INTENSIDAD_SINTOMATICA,u.EMBARAZO_CONTROLADO,u.CEFALEAS,u.EDEMA
FROM consulta_external AS c
JOIN descripcionqx1 AS d
ON (c.id=d.id)
JOIN epicrisis1 AS e
ON (d.id=e.id)
JOIN parto1 AS p
ON (e.id=p.id)
JOIN urgencial AS u
ON (p.id=u.id)
ORDER BY id DESC;

```

ID	EDAD	ESTRATO	CODIGO_DIAGNOSTICO_PRINCIPAL	CODIGO_DIAGNOSTICO_PRINCIPAL_SALIDA	CODIGO_DIAGNOSTICO_REL	
1	99986	26	1	O471	Z392	O420
2	99973	20	1	O623	O839	O429
3	99917	17	1	O034	O021	R102
4	99917	17	1	O034	O021	R102

Ilustración 11 Código SQL – Fuente propia

Una vez teniendo los datos organizados en una sola tabla, se procede a crear una nueva tabla con la Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud, Decima Revisión (CIE-10), para identificar a qué tipo de diagnóstico pertenece cada uno de los casos relacionados en el proyecto.

Para la creación de esta tabla se ejecutará el siguiente código como lo muestra la imagen el cual contará con dos campos *código* y *descripción* de tipo “*varchar*”, en el que el campo *código* obtendrá el numero identificador del diagnóstico y *descripción* obtendrá el nombre de ese diagnóstico. (Ver ilustración 12)

```

USE [PROYECTO]
GO

/***** Object: Table [dbo].[CIE10]    Script Date: 27/09/2019 19:00:22 *****/
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

SET ANSI_PADDING ON
GO

CREATE TABLE [dbo].[CIE10](
    [CODIGO] [varchar](50) NULL,
    [DESCRIPCION] [varchar](200) NULL
) ON [PRIMARY]
GO

SET ANSI_PADDING OFF
GO

```

Ilustración 12 Código SQL – Fuente propia

Creada la tabla en mención anterior, se realiza posteriormente la inserción de datos a la tabla CIE10, como lo muestra la ilustración 13.

```

INSERT INTO dbo.[CIE10]
    ([CODIGO]
    ,[DESCRIPCION])
VALUES
    ('A000', 'COLERA DEBIDO A VIBRIO CHOLERAEE 01, BIOTIPO CHOLERAEE')
GO
INSERT INTO dbo.[CIE10]
    ([CODIGO]
    ,[DESCRIPCION])
VALUES
    ('A001', 'COLERA DEBIDO A VIBRIO CHOLERAEE 01, BIOTIPO EL TOR')
GO
INSERT INTO dbo.[CIE10]
    ([CODIGO]
    ,[DESCRIPCION])
VALUES
    ('A009', 'COLERA NO ESPECIFICADO')
GO
INSERT INTO dbo.[CIE10]
    ([CODIGO]
    ,[DESCRIPCION])
VALUES
    ('A010', 'FIEBRE TIFOIDEA')
GO
INSERT INTO dbo.[CIE10]
    ([CODIGO]
    ,[DESCRIPCION])
VALUES
    ('A011', 'FIEBRE PARATIFOIDEA A')
GO

```

Ilustración 13 Código SQL – Fuente propia

Como resultado de la creación e inserción de datos de la tabla “CIE10” (ver ilustración 14):

	CODIGO	DESCRIPCION
1	A000	COLERA DEBIDO A VIBRIO CHOLERAЕ O1, BIOTIPO CHOLE...
2	A001	COLERA DEBIDO A VIBRIO CHOLERAЕ O1, BIOTIPO EL TOR
3	A009	COLERA NO ESPECIFICADO
4	A010	FIEBRE TIFOIDEA
5	A011	FIEBRE PARATIFOIDEA A
6	A012	FIEBRE PARATIFOIDEA B
7	A013	FIEBRE PARATIFOIDEA C
8	A014	FIEBRE PARATIFOIDEA, NO ESPECIFICADA
9	A020	ENTERITIS DEBIDA A SALMONELLA
10	A021	SEPTICEMIA DEBIDA A SALMONELLA
11	A022	INFECCIONES LOCALIZADAS DEBIDA A SALMONELLA

Ilustración 14 Código SQL – Fuente propia

Creada la tabla CIE-10 procedemos a crear los subgrupos que categorizan los criterios de inclusión de la MME.(Ver ilustración 15)

```
USE PROYECTO
GO

CREATE TABLE GRUPOS (
  CODIGO varchar (500),
  DESCRIPCION varchar (500),
);

USE PROYECTO
GO

CREATE TABLE GRUPOS_DIAGNOSTICOS (
  CODIGO_GRUPO varchar(500),
  CODIGO_DX varchar(500),
);
```

Ilustración 15 Código SQL – Fuente propia

Posteriormente a la creación de las tablas de los grupos y subgrupos de los criterios de inclusión materna teniendo en cuenta el *CIE-10* se procede a realizar la inserción de los mismo de acuerdo con su clasificación dependiendo del grupo que pertenece con sus respectivos grupos diagnósticos como se aprecia en las ilustraciones 16 y 17.

```
USE [PROYECTO]
GO

INSERT INTO [dbo].[GRUPOS]([CODIGO],[DESCRIPCION])
VALUES ('B99','Otras enfermedades infecciosas')
GO
```

Ilustración 16 Código SQL – Fuente propia

```
USE [PROYECTO]
GO

INSERT INTO [dbo].[GRUPOS_DIAGNOSTICOS]
([CODIGO_GRUPO],[CODIGO_DX])
SELECT 'B99X',CODIGO FROM CIE10 WHERE CODIGO ='B99X' ORDER BY CODIGO
GO
```

Ilustración 17 Código SQL – Fuente propia

A través de este script (ver ilustración 18) se centralizan los grupos con los criterios de inclusión en una sola tabla para su posterior manejo.

```

USE PROYECTO
GO

select R.GRUPO, COUNT(R.GRUPO)VECES from (SELECT
CASE when C.codigo between 'A00' and 'A09x' then 'A00-A09'
when C.codigo between 'A15' and 'A19x' then 'A15-A19'
when C.codigo between 'A20' and 'A28x' then 'A20-A28'
when C.codigo between 'A30' and 'A49x' then 'A30-A49'
when C.codigo between 'A50' and 'A64x' then 'A50-A64'
when C.codigo between 'A65' and 'A69x' then 'A65-A69'
when C.codigo between 'A70' and 'A74x' then 'A70-A74'
when C.codigo between 'A75' and 'A79x' then 'A75-A79'
when C.codigo between 'A80' and 'A89x' then 'A80-A89'
when C.codigo between 'A90' and 'A99x' then 'A90-A99'
when C.codigo between 'B00' and 'B09x' then 'B00-B09'
when C.codigo between 'B15' and 'B19x' then 'B15-B19'
when C.codigo between 'B20' and 'B24x' then 'B20-B24'
when C.codigo between 'B25' and 'B34x' then 'B25-B34'
when C.codigo between 'B35' and 'B49x' then 'B35-B49'
when C.codigo between 'B90' and 'B94x' then 'B90-B94'
when C.codigo between 'B50' and 'B64x' then 'B50-B64'
when C.codigo between 'B65' and 'B83x' then 'B65-B83'
when C.codigo between 'B65' and 'B83x' then 'B65-B83'
when C.codigo between 'B85' and 'B89x' then 'B85-B89'
when C.codigo between 'B95' and 'B97x' then 'B95-B97'
when C.codigo between 'B99' and 'B99x' then 'B99'
when C.codigo between 'R75' and 'R75x' then 'R75'

```

Ilustración 18 Código SQL – Fuente propia

Posteriormente procedemos agregar los diagnósticos de inclusión faltantes para un mejor procesamiento de los datos en el siguiente script de la base de datos como lo muestra la ilustración 19.

```

SELECT *
into PACIENTES_MME
FROM epicrisis WHERE ID IN (SELECT ID FROM [dbo].[consulta_externa]) AND
(CodigoDiagnosticoPrincipal IN ('0140','0141','0149','0150','0151','0152','0159','016X','085X','0468','0469','0720'
OR CodigoDiagnosticoRelacionado1 IN ('0140','0141','0149','0150','0151','0152','0159','016X','085X','0468','0469','
OR CodigoDiagnosticoRelacionado2 IN ('0140','0141','0149','0150','0151','0152','0159','016X','085X','0468','0469','
OR CodigoDiagnosticoRelacionado3 IN ('0140','0141','0149','0150','0151','0152','0159','016X','085X','0468','0469','
OR CodigoDiagnosticoRelacionado1Salida IN ('0140','0141','0149','0150','0151','0152','0159','016X','085X','0468','C
OR CodigoDiagnosticoRelacionado2Salida IN ('0140','0141','0149','0150','0151','0152','0159','016X','085X','0468','C
OR CodigoDiagnosticoRelacionado3Salida IN ('0140','0141','0149','0150','0151','0152','0159','016X','085X','0468','C

```

Ilustración 19 Código SQL – Fuente propia

Finalmente, cuando se han realizado los respectivos filtros de las variables quedan los justamente necesarios para la implementación de la red neuronal y realizar las respectivas pruebas de entrenamiento y procesamientos de los datos en el algoritmo. (Ver ilustración 20 e ilustración 21).

	ID	EDAD	ESTRATO	EDAD_MADRE	OBSTETRICOS_GESTACIONES	OBSTETRICOS_PARTOS	OBSTETRICOS_ABORTOS
1	157603	28	1	27	3	3	0
2	157613	17	1	16	1	1	0
3	157781	27	1	27	3	2	1
4	157799	17	1	16	2	1	1
5	157802	16	1	15	1	1	0
6	157958	36	1	36	4	4	0
7	157831	29	1	28	2	2	0
8	157910	36	1	36	7	07	0
9	157922	30	1	29	3	2	1
10	157933	35	1	34	6	4	2

Ilustración 20 Código SQL – Fuente propia

OBSTETRICOS_CESAREAS	MESES_SEMANAS	GRUPO_RH_MADRE	CP	VIVO_MUERTO	SEXO	O00-O08	O10-O97	O30-O48
0	382	1	1	1	1	0	0	0
1		17	1	1	1	0	0	0
0	402	1	1	1	1	0	0	0
1	384	1	1	1	0	0	0	0
0	382	1	1	1	0	0	0	0
0	402	1	1	1	0	0	0	0
0	40	1	1	1	0	0	0	0
0		1	0	1	0	0	0	0
2	382	17	1	1	1	0	0	0
0		1	1	1	0	0	0	0

Ilustración 21 Código SQL – Fuente propia

3.1.2. Análisis de las variables de riesgo

El conjunto de datos fue entregado con un total de 26.231 datos y 68 columnas, como se puede observar en la ilustración 30, 31 y 32. En este Dataset se encontraron dos columnas irrelevantes como la columna ID y en el caso de (meses_semanas) porque posee una gran cantidad de valores nulos y mediante las funciones dropna se eliminarán, como se puede observar en la ilustración 22. Por medio de las siguientes funciones *np.array*, *isnull()*, *info()*, *missing_values_count*, se realizará un análisis para seleccionar las variables dependientes e independientes del conjunto de datos.

Selección de variable dependiente o Labels:

La última columna del Dataset llamada *MME* es la columna seleccionada para ser la variable dependiente o variable de resultado. Por medio de la función *np.array* de numpy, será usada como variable dependiente por el algoritmo para generar un modelo que prediga lo mejor posible si una paciente en medio de su proceso de embarazo pueda llegar a padecer algún criterio de inclusión y ser clasificada o no como Morbilidad Materna Extrema.

Columna con valores nulos columna (MESES SEMANAS)

Con la librería Pandas *Dataframe* proporciona una función *isnull ()*, la cual devuelve un nuevo marco de datos del mismo tamaño que el marco de datos original, contiene solo verdadero (True) y falso (False). Con True en el lugar NaN o campos nulos en el marco de datos y False en otros lugares.

Columna MESES_SEMANAS.

OBSTETRICOS_CESAREAS	MESES_SEMANAS	GRUPO_RH_MADRE	CP
False	False	False	False
False	True	False	False
False	False	False	False
False	False	False	False
False	False	False	False
...
False	False	False	False
False	True	False	False
False	True	False	False
False	False	False	False
False	False	False	False

Ilustración 22 Valores nulos en la columna MESES_SEMANAS

Columna MESES_SEMANAS del conjunto de datos, por medio de la función *info* de la Librería Pandas y se usa para obtener un resumen conciso del marco de datos, resulta muy útil al hacer un análisis exploratorio de los datos.

```

MESES_SEMANAS
0      382.0
1       NaN
2      402.0
3      384.0
4      382.0
...      ...
26226   376.0
26227   NaN
26228   NaN
26229    39.0
26230   403.0

```

Ilustración 23 Columna MESES_SEMANAS con datos nulos

Por medio de la función *missing_values_count*, se visualiza la cantidad de datos nulos que hay en el conjunto de datos.


```
[30] missing_values_count[:10]
```

```
↳ ID                0
   EDAD              0
   ESTRATO           0
   EDAD_MADRE        0
   OBSTETRICOS_GESTACIONES 0
   OBSTETRICOS_PARTOS 0
   OBSTETRICO_ABORTOS 0
   OBSTETRICOS_CESAREAS 0
   MESES_SEMANAS     9579
   GRUPO_RH_MADRE    0
dtype: int64
```

Ilustración 24 Cantidad de datos nulos

Por medio de la función *dropna* se busca eliminar la columna *MESES_SEMANAS* del conjunto de datos. Esta columna cuenta con una cantidad de 9.579 datos nulos, los cuales generan ruido en el proceso de generalización de la información.

```
Index(['ID', 'EDAD', 'ESTRATO', 'EDAD_MADRE', 'OBSTETRICOS_GESTACIONES',
      'OBSTETRICOS_PARTOS', 'OBSTETRICO_ABORTOS', 'OBSTETRICOS_CESAREAS',
      'GRUPO_RH_MADRE', 'CP', 'VIVO_MUERTO', 'SEXO', '000-008', '010-097',
      '030-048', '060-092', 'P01-P03', 'P01-P03.1', 'P00-P04', 'P05-P07',
      'P10-P15', 'P20-P28', 'P35-P39', 'Q20-Q28', 'Q90-Q99', 'E10-E14',
      'Q10-Q89', 'Z00-Z99', 'R00-R96', 'N60-N98', 'D10-D36', 'N30-N39',
      'D50-D64', 'E00-E90', 'K00-K93', 'P20-P28.1', 'J00-J99', 'P05-P07.1',
      'G04-H95', 'A00-A09', 'I10-I15', 'N00-N29', 'J45-J46', 'P01-P03.2',
      'A50-A64', 'L00-L99', 'P35-P39.1', 'D37-D45', 'Q10-Q89.1', 'E10-E14.1',
      'D65-D77', 'A30-A49', 'Y40-Y84', 'F17-F99', 'K50-K52', 'B00-B09',
      'B20-B24', 'C53', 'J12-J18', 'B95-B97', 'A65-A69', 'I71-I99', 'K80-K83',
      'B35-B49', 'A15-A19', 'D00-D09', 'MME'],
      dtype='object')
```

Ilustración 25 Conjunto de datos, sin la columna MESES_SEMANAS

Cantidad de 0 y 1 respecto a la variable dependiente o de salida (MME).

```
print(datos_.groupby('MME').count())
```

	ID	EDAD	ESTRATO	EDAD_MADRE	...	K80-K83	B35-B49	A15-A19	D00-D09
MME					...				
0	22875	22875	22875	22875	...	22875	22875	22875	22875
1	3356	3356	3356	3356	...	3356	3356	3356	3356

[2 rows x 67 columns]

Ilustración 26 Cantidad de 0 y 1, relación a la variable de salida

En la ilustración 28 se observa la cantidad de datos que se utiliza para el entrenamiento, mostrando la cantidad de 1 y 0 que se encuentran en la variable de salida *y_train*. Arrojando como resultado 3.356 datos con valor uno (1) y 22.875 con valor cero (0) para las salidas de entrenamiento.

```
# Dataset
X_train = X_train[:]
y_train = y_train[:]

print(sum(y_train==1))
print(sum(y_train==0))

[959]
[6041]
```

Ilustración 27 - 0 y 1 (entrenamiento)

Con el método *hist* podemos observar en el eje Y la cantidad de datos y en el eje X el rango de edades de las pacientes.

<matplotlib.axes._subplots.AxesSubplot at 0x7f3d8a8f6160>

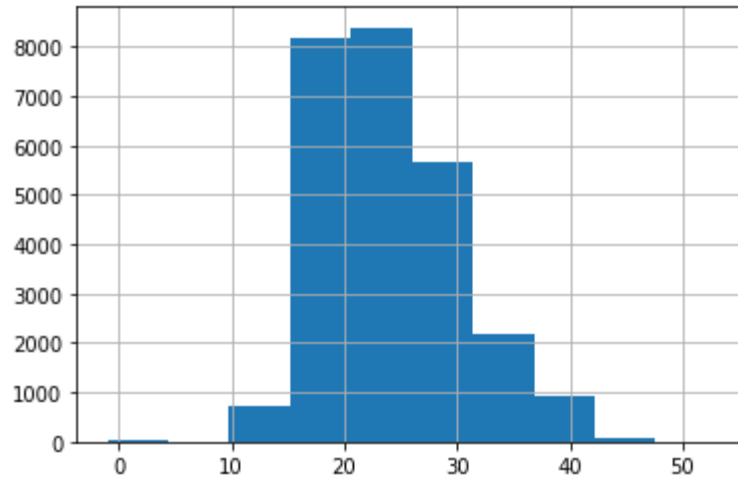


Ilustración 28 Rango de edades relación al conjunto de datos

3.1.3. Construcción del conjunto de datos

Seguido a todo el proceso de la obtención y filtrado de los datos extraídos de la base de datos se exportan a un archivo .csv, el cual mediante la implementación de librerías propias utilizadas por el lenguaje de programación Python, permita al algoritmo realizar la predicción de los casos de MME.

Inicialmente se importa la librería de Pandas y luego por medio de la función `read_csv`, se procede a cargar el conjunto de datos (*Dataset*) para lograr visualizarlos y tener una mejor comprensión de los datos al momento de trabajar con ellos.

Este conjunto de datos cuenta con:

Cantidad de filas	26.231
Cantidad de columnas	68

Columnas del Dataset:

ID	EDAD	ESTRATO	EDAD_MADRE	OBSTETRICOS_GESTACIONES	OBSTETRICOS_PARTOS	OBSTETRICO_ABORTOS	OBSTETRICOS_CESAREAS	MESES_SEMANAS	GRUPO_RH_MADRE	CP	VIVO_MUERTO	S
0	157603	28	1	27		3	3	0	0	382.0	1 1	1
1	157613	17	1	16		1	1	0	1	NaN	17 1	1
2	157781	27	1	27		3	2	1	0	402.0	1 1	1
3	157799	17	1	16		2	1	1	1	384.0	1 1	1
4	157802	16	1	15		1	1	0	0	382.0	1 1	1
...
26226	157469	21	1	20		1	1	0	1	376.0	1 1	1
26227	157483	17	1	16		1	1	0	0	NaN	1 1	1
26228	157483	17	1	16		1	1	0	0	NaN	1 1	1
26229	157567	32	1	32		2	2	0	2	39.0	17 1	1
26230	157579	23	1	23		2	2	0	0	403.0	1 1	1

26231 rows × 68 columns

Ilustración 29 Conjunto de datos – fuente propia

SEXO	000-008	010-097	030-048	060-092	P01-P03	P01-P03.1	P00-P04	P05-P07	P10-P15	P20-P28	P35-P39	Q20-Q28	Q90-Q99	E10-E14	Q10-Q09	Z00-Z99	R00-R96	N60-N98	D10-D36	N30-N39	D50-D64	E00-E90	K00-K93	P20-P28.1	J00-J99	P05-P07.1	G04-G09	A00-A09	I10-I15	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
...
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Ilustración 30 Conjunto de datos – fuente propia

N00-N29	J45-J46	P01-P03.2	A50-A64	L00-L99	P35-P39.1	D37-D45	Q10-Q09.1	E10-E14.1	D65-D77	A30-A49	Y40-Y84	F17-F99	K50-K52	B00-B09	B20-B24	C53	J12-J18	B95-B97	A65-A69	I71-I99	K80-K83	B35-B49	A15-A19	D00-D09	MME		
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
...
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Ilustración 31 Conjunto de datos - fuente propia

El conjunto de datos fue entregado con un total de 26.231 datos y 68 columnas, como se puede observar en la ilustración 1, 2 y 3.

4. MODELADO DE LA RED NEURONAL

El presente capítulo describe la construcción, el desarrollo y la evaluación de la red neuronal profunda (Deep Learning) para la predicción de riesgos de Morbilidad Materna Extrema (MME) Por medio de librerías como Keras de Python, utiliza funciones como *model.compile()*, *model.fit()*, *model.evaluate()* se logra compilar, entrenar y evaluar el modelo de clasificación usando la API de Keras para la elaboración de algoritmos de Machine Learning desarrollada en Python y capaz de ejecutarse sobre TensorFlow.

Así mismo se estudiarán y aplicarán las técnicas de filtrado de variables, normalización de los datos, división del conjunto de datos para entrenamiento y prueba, el balanceo de los datos y la función *one-hot encode*.

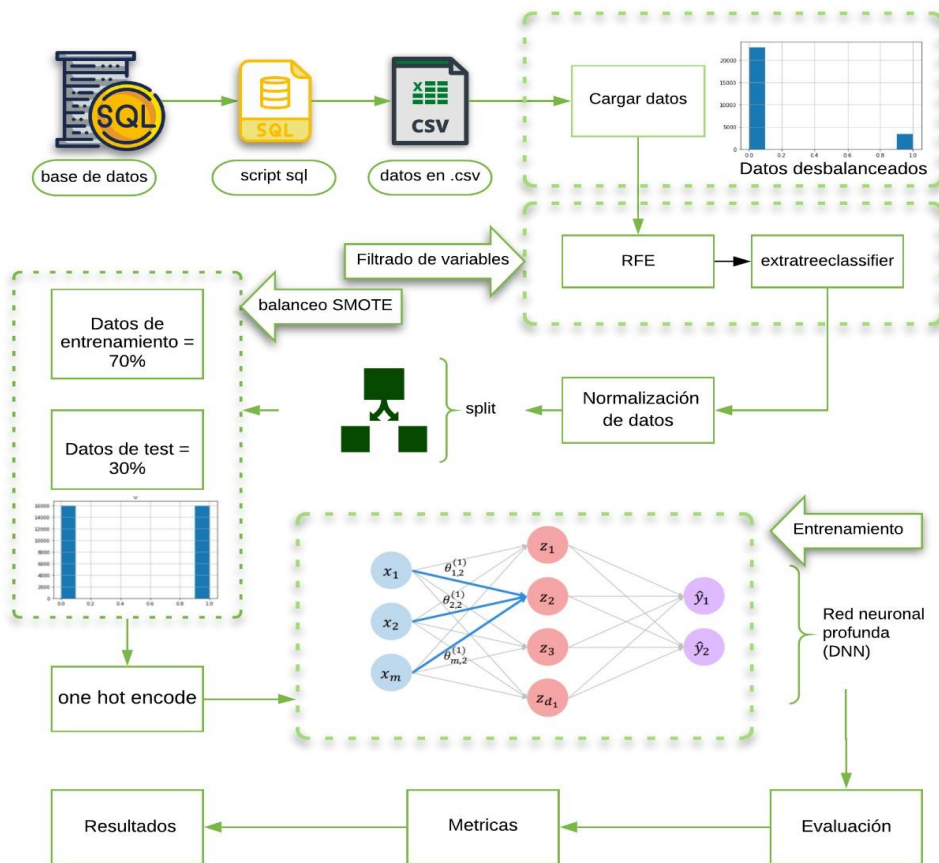


Ilustración 32 Diseño del Algoritmo

4.1. Técnicas de filtrado de variables

En este modelo fue necesario aplicar técnicas de filtrado de variables como *RFE Recursive Feature Elimination* (Eliminación de características recursivas), un tipo de filtrado que busca adaptarse al modelo y eliminar las características más débiles según la selección de los datos que se realiza por medio del clasificador *DecisionTreeClassifier()*.

```
filtrado_RFE(datos, labels_)  
  
['EDAD' 'ESTRATO' 'EDAD_MADRE' 'OBSTETRICOS_GESTACIONES'  
'OBSTETRICOS_PARTOS' 'OBSTETRICO_ABORTOS' 'OBSTETRICOS_CESAREAS'  
'GRUPO_RH_MADRE' 'CP' 'VIVO_MUERTO' 'SEXO' 'O10-097' 'O30-048' 'O60-092'  
'P01-P03' 'P00-P04' 'P05-P07' 'P10-P15' 'P20-P28' 'P35-P39' 'Q10-Q89'  
'Z00-Z99' 'P20-P28.1' 'J00-J99' 'P05-P07.1' 'G04-H95' 'P01-P03.2'  
'A50-A64' 'P35-P39.1' 'Q10-Q89.1' 'D65-D77' 'B20-B24']
```

Ilustración 33 Resultado de la técnica de filtrado RFE

Se obtuvo como resultado del filtrado de variables *RFE* y sus parámetros, un nuevo grupo de columnas de las cuales se seleccionarán las 20 más relevantes para el entrenamiento del modelo. Por medio del filtrado *ExtraTreesClassifier()* se implementa un metaestimulador que ajusta varios árboles de decisión aleatorios en varias submuestras del conjunto de datos, utilizando el promedio para mejorar la precisión predictiva y controlar el sobreajuste (overfitting).

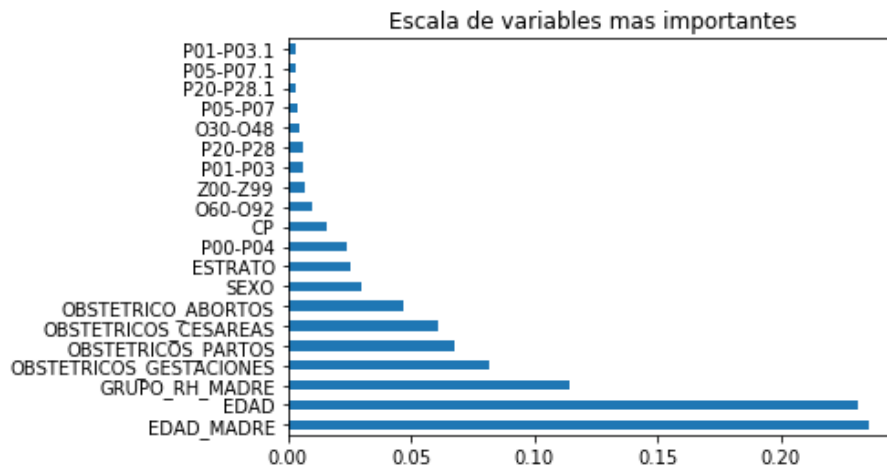


Ilustración 34 Resultado de la técnica de filtrado

Como resultado se obtuvieron las 20 columnas con características significativas para el filtrado de datos, la columna EDAD_MADRE ubicada como la más significativa.

4.2. Normalización de los datos

A través de la función *MinMaxScaler()* transforma las características a un rango dado, por defecto entre (0 y 1) aunque puede ser personalizado, solo interesa escalar las características predictivas, de forma que entrena el escalador y con la función *fit.transform()* transforma las columnas arrojadas por el algoritmo de filtrado *RFE* en un array bidimensional.

```
[[0.53703704 0.          0.01796821 ... 0.          0.          0.          ]
 [0.33333333 0.          0.01036628 ... 0.          0.          0.          ]
 [0.51851852 0.          0.01796821 ... 0.          0.          0.          ]
 ...
 [0.33333333 0.          0.01036628 ... 0.          0.          0.          ]
 [0.61111111 0.          0.02142364 ... 0.          0.          0.          ]
 [0.44444444 0.          0.01520387 ... 0.          0.          0.          ]
 (26231, 20)
```

Ilustración 35 Resultado de la normalización

Como resultado de la normalización se obtuvo un array con rango de (0-1) con una cantidad de 26.231 filas y 20 columnas elegidas por el filtrado *RFE*.

4.3. División de los datos, entrenamiento y prueba

En todo modelo de aprendizaje supervisado es necesario dividir el conjunto de datos en dos partes una destinada a entrenamiento (*train*) 70% y otra parte destinada a prueba (*test*) 30. Usando la función (*train_test_split()*) se realiza la división del conjunto de datos y con el parámetro *random_state* se establece las cantidades de datos en forma aleatoria.

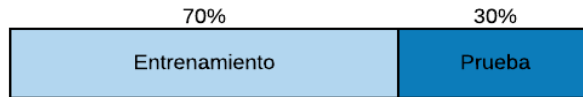


Ilustración 36 Split de entrenamiento y prueba - (Fuente propia)

	Filas	Columnas
x_train	18.361	20
y_train	18.361	1
x_test	7.870	20
y_test	7.870	1
Total, de los datos 26.231		

Tabla 7 Resultado de las cantidades de filas y columnas en la división del Dataset – Fuente propia

4.4. Balanceo de los datos

Los datos desbalanceados afectan al modelo en su proceso de generalización de la información por lo cual perjudica a las clases minoritarias. Para combatir este desbalance en los datos se utilizó la técnica *smote* de sobremuestreo de la minoría de clases, en este caso la cantidad de 0 era superior a las cantidades de 1, como se observa en la ilustración 37.

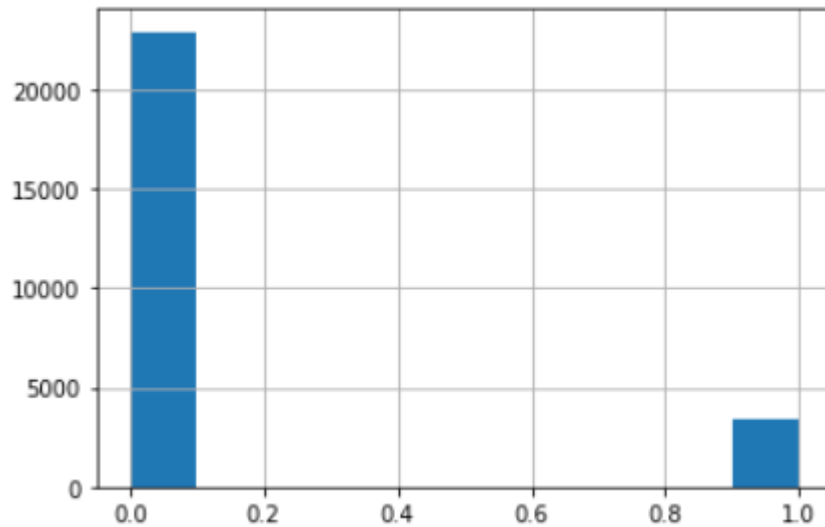


Ilustración 37 Datos desbalanceados

En la tabla 7 se muestra la cantidad de datos de la clase '0' y '1' antes y después del balanceo:

Técnica de sobremuestreo SMOTE	Cantidad de datos	Columnas
Antes del sobremuestreo, conteo de la clase '1'	960	
Antes del sobremuestreo, conteo de la clase '0'	6.910	
Después del sobremuestreo, la forma de train_X	13.820	20
Después del sobremuestreo, la forma de train_y	13.820	1
Después del sobremuestreo, conteo de la clase '1'	6.910	
Después del sobremuestreo, conteo de la clase '0'	6.910	

Tabla 8 Resultado de la técnica smote, balanceando la cantidad de 0 y 1

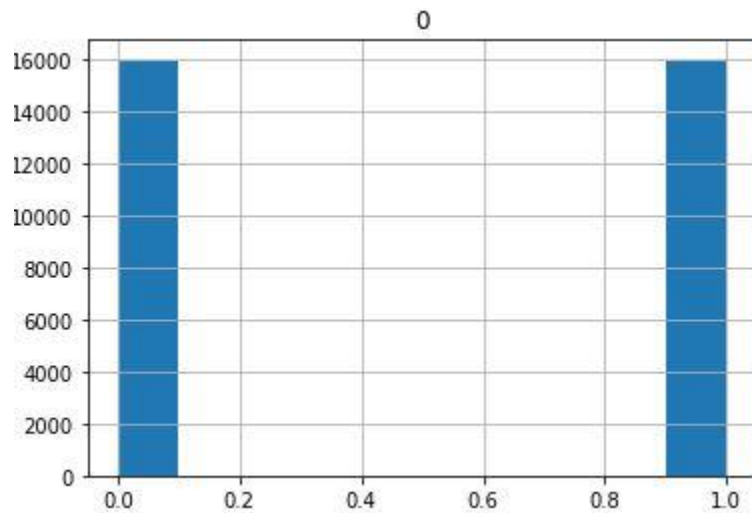


Ilustración 38 Resultado de la clase 0 y 1 balanceadas

Como resultado se obtuvo una cantidad de 31.930 registros de los cuales se dividen en 2 partes, 15.965 para la cantidad de '0' y 15.965 para la cantidad de '1' obteniendo un conjunto de datos balanceados.

Los datos de pruebas también tuvieron que ser balanceados ya que para la cantidad de '0' tenían 6.910 valores y la cantidad de '1' 960 valores.

Dataset	Filas	Columnas
x_train	31930	20
y_train	31930	2
x_test	13.820	20
y_test	13.820	2

Tabla 9 Resultado con la cantidad de filas y columnas balanceadas en el Dataset - fuente propia

Como resultado se obtuvieron 31.930 datos para entrenamiento y 13.820 para prueba con 20 columnas y 2 clases de salidas '0 no MME' y '1 si MME'

One-hot encode (Codificación activa)

Lo que hace una codificación activa es que toma una columna que tiene datos categóricos, que ha sido codificada en etiquetas, y luego divide la columna en varias columnas. Los números se reemplazan por 1s y 0s, dependiendo de qué columna tenga qué valor.

Dataset	Filas	Columnas
x_train	31930	20
y_train	31930	2
x_test	13.820	20
y_test	13.820	2

Tabla 10 Resultado de la visión de la variable dependiente MME

Como resultado arroja dos columnas dependientes, '0' que no es clasificada con MME y '1' una paciente que si es clasificada con MME.

4.5. Construcción del algoritmo de predicción

En la construcción del modelo se han definido 2 capas densamente conectadas, es decir, que todas las neuronas de cada capa están conectadas con todas las

neuronas de la capa siguiente, a esta primera capa densa (*dense*) se le asigna 128 neuronas, un `kernel_initializer` normal que define la forma de establecer los pesos, una función de activación Relu (ya que arroja mejores resultados que la sigmoide) y un `Bias_initializer` ones para generar tensores inicializados en 1 (ver tabla 10).

Estructura del modelo de red neuronal				
tf.keras.models.Sequential()				
Agregar capas	Tipo de capas	Kernel_initializer	Función de activación	Bias_initializer
.add	Dense	normal	relu	ones
.add	Dropout			
.add	Dense	normal	softmax	ones

Tabla 11 Construcción de la red neuronal

A la segunda capa se le añade una regularización de tipo (*Dropout*) de (0,5), la regularización lo que hace es apagar el 50% de las neuronas de forma aleatoria. A la tercera capa se le agregan 2 neuronas, con `kernel_initializer` normal y una función de activación softmax y un `Bias_initializer` ones para generar tensores inicializados en 1.

Penalización de los pesos

En este modelo se tuvo que utilizar un parámetro adicional (`class_weights`) para la estimación de los pesos para las clases del conjunto de datos.

Donde (`class_weights = {0: 18.0, 1: 1}`) donde se fuerza al algoritmo a tratar cada instancia de la clase '0' no MME se le asigna una penalización de 18 instancias, igual que a la clase '1' si es MME.

Compilación del modelo

Compilación del modelo			
model.compile()			
Optimizador	Perdida	Métricas	Learning Rate
adam	categorical_crossentropy	exactitud	0.001

Tabla 12 Parámetros para la compilación del modelo

Para la compilación del modelo fue necesario probar diferentes optimizadores, el que más se ajustó al modelo fue el optimizador adam encargado de actualizar los pesos de red de forma iterativa en función de los datos de entrenamiento. La función de costo seleccionada fue la *categorical_crossentropy* la cual mide el rendimiento de un modelo de clasificación cuyo resultado es un valor de probabilidad entre 0 y 1. La métrica seleccionada para monitorear los pasos de entrenamiento y prueba es *accuracy* (exactitud), la fracción de los datos que son correctamente clasificados. El *learning rate* (tasa de aprendizaje) es el método que regula la velocidad para encontrar el mínimo local de error.

4.6. Entrenamiento del modelo

Para la etapa de entrenamiento fue necesario agregar el método *model.fit* para el entrenamiento del modelo con un número fijo de épocas.

Entrenamiento del modelo			
model.fit()			
Batch_size	Validation_data	Shuffle	Epochs
32	X_test_res ; y_test	True	800

Tabla 13 Parámetros para el entrenamiento del modelo

Se deben ajustar los parámetros como el *batch_size* con un valor de 32 ya que van a hacer 32 muestras que se propagarán a través de la red neuronal.

Así mismo *validation_data* ayuda a seleccionar los datos sobre los cuales va a evaluar la pérdida y cualquier métrica del modelo al final de cada época.

Por otra parte, barajar (*shuffle=True*) baraja los datos de entrenamiento antes de cada época y El número de épocas (*epochs*) son 800 épocas para entrenar el modelo, realiza una iteración sobre todo *X_train* y *y_train* de datos proporcionados.

Resultados del con el bloque de entrenamiento			
Perdida	Exactitud	Validación de pérdida	Validación de exactitud
10.77%	0.64%	0.68%	0.64%

Tabla 14 Resultados del entrenamiento del modelo

4.7. Evaluación del modelo

Una vez la red neuronal se encuentra entrenada, puede ser evaluada mediante la función *evaluate()* pasándole el conjunto de datos no visto durante el entrenamiento, ese 30% de datos de prueba separado anteriormente proporciona una estimación del rendimiento de la red a la hora de hacer predicciones de datos no vistos en el futuro. El modelo evalúa la pérdida a través de todos los patrones de prueba, así como cualquier otra métrica especificada cuando se compiló el modelo como la exactitud.

Evaluación del modelo	
model.evaluate()	
Precisión	Perdida
0.64%	0.69%

Tabla 15 Resultado con base a la precisión y perdida

Graficas del margen de exactitud y perdida

Por medio de graficas se puede visualizar mejor el margen de exactitud (*acc*) y perdida *loss* recopilado del historial de entrenamiento y validación. Usado mucho en modelos de aprendizaje profundo (*Deep Learning*), la función utilizada es *history*, es la que registra el comportamiento de las métricas aplicadas en entrenamiento y validación para cada época del modelo clasificadorio.

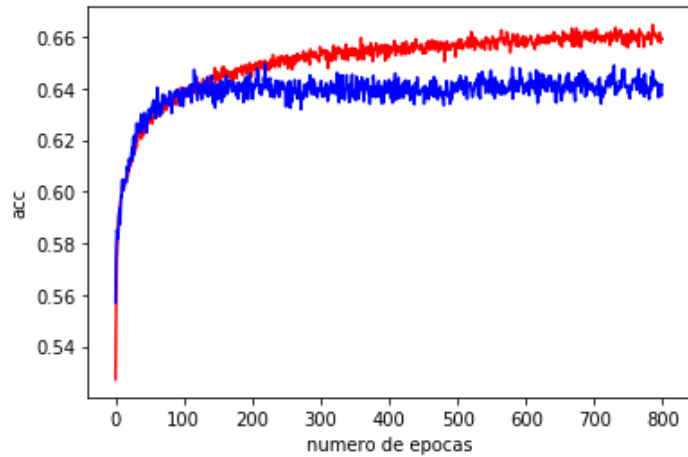


Ilustración 39 Resultado exactitud y perdida ; entrenamiento (*train*) vs validación (*test*)

Esta grafica representa el resultado del margen de exactitud del entrenamiento (*train*) vs validación (*test*). En color rojo: se muestra el entrenamiento (*train*), y en color azul el resultado pruebas (*test*)

Perdida (loss)

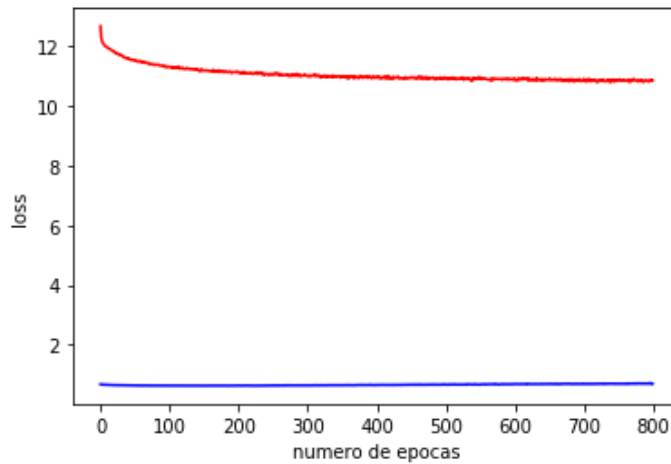


Tabla 16 Grafica del margen de error ; entrenamiento (*train*) vs validación (*test*)

La gráfica anterior se interpreta más fácilmente en la tabla 16, donde se muestra el resultado de la exactitud y la pérdida, tanto en entrenamiento como en pruebas.

Resultado	Evaluación de entrenamiento	Evaluación de prueba
Perdida	10.77%	0.69%
Exactitud	0.66%	0.64%
Validación perdida	0.68%	
Validación exactitud	0.64%	

Tabla 17 Resultado de la evaluación del modelo - train vs test

Métricas de evaluación del modelo

La **matriz de confusión** es una de las métricas más intuitivas y sencillas que se utilizan para encontrar la precisión y exactitud del modelo. Es una matriz que describe el rendimiento del modelo de clasificación con los datos de prueba y donde las salidas pueden ser dos o más clases.

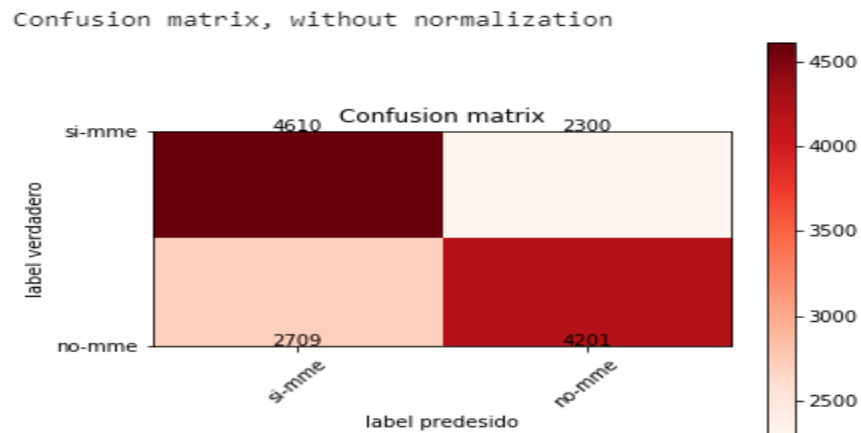


Ilustración 40 Resultado en la matriz de confusión

Donde,

- VP representa los verdaderos positivos que corresponden a los pacientes con MME (1) en la vida real vs Los pacientes con MME (1) según la predicción del modelo.
- VN corresponde a los verdaderos negativos, que son los pacientes que realmente NO tiene MME (0) y el modelo predice que NO tiene MME (0).

- FP o falsos positivos, son pacientes que NO tiene MME realmente (0) y el modelo predice que si tiene MME (1).
- FN, son los falsos negativos, es decir, pacientes que tiene MME en la vida real (1) y que el modelo lo predice como negativo para MME (0).

Como resultado del análisis de la matriz de confusión, se obtiene la tabla 17 que muestra el total de predicciones hechas por el modelo para cada término de la matriz.

Exactitud (accuracy)

Porcentaje total de los aciertos de nuestro modelo.

$$\text{Formula: } \frac{4.610+4.201}{4.610+2.300+2.709+4.201} = 0.64\% \text{ de exactitud}$$

Precisión (tasa de error)

Llamado valor de predicción positiva. Nos da la probabilidad de que, dada una predicción positiva, la realidad sea positiva también.

$$\text{Formula: } \frac{4.610}{4.610+2.300} = 0.67\% \text{ de precisión}$$

Sensibilidad (recall)

Tasa de verdaderos positivos. Nos da la probabilidad de que, dada una observación realmente positiva, el modelo la clasifique así.

$$\text{Formula: } \frac{4.610}{4.610+2.709} = 0.63\% \text{ de sensibilidad}$$

Medida F (F1)

Esta medida se considera una medida armónica entre la sensibilidad y la precisión.

$$\text{Formula: } \frac{2*0.67*0.63}{0.67+0.63} = 0.65\% \text{ de F1}$$

Prevalencia

Es la probabilidad de un positivo en el total de la muestra.

$$\text{Formula: } \frac{4.610+2.709}{4.610+4.201+2.300+2.709} = 0.53\% \text{ de prevalencia}$$

Error de clasificación

Porcentaje de errores del modelo.

$$\text{Formula: } \frac{2.300+2.709}{4.610+4.201+2.300+2.709} = 0.36\% \text{ de error de clasificación}$$

Clasificación ROC y AUC

Una curva ROC (curva de característica operativa del recepto) es un gráfico que muestra el rendimiento de un modelo de clasificación en todos los umbrales de clasificación. Esta curva representa dos parámetros:

- Tasa de verdaderos positivos
- Tasa de falsos positivos

El AUC es el área bajo la curva ROC. Este puntaje nos da una buena idea de qué tan bien funciona el modelo.

- Es conveniente por las siguientes razones:
- Es invariable con respecto a la escala, mide qué tan bien se clasifican las predicciones, en lugar de sus valores absolutos.
- Es invariable con respecto al umbral de clasificación, mide la calidad de las predicciones del modelo, sin tener en cuenta qué umbral de clasificación se elige.

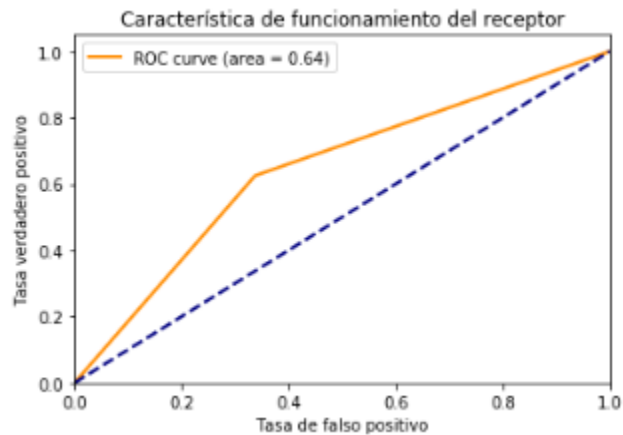


Ilustración 41 Curva ROC ; VP y FP

- Tasa de verdaderos positivos (TPR)

formula: $\frac{VP}{VP+FN}$; $\frac{4481}{4481+2429} = 0,64\%$ de sensibilidad

- Tasa de falsos positivos (FPR)

formula: $\frac{FP}{FP+VN}$; $\frac{2527}{4481+4383} = 0,28\%$ de especificidad

5. RESULTADOS

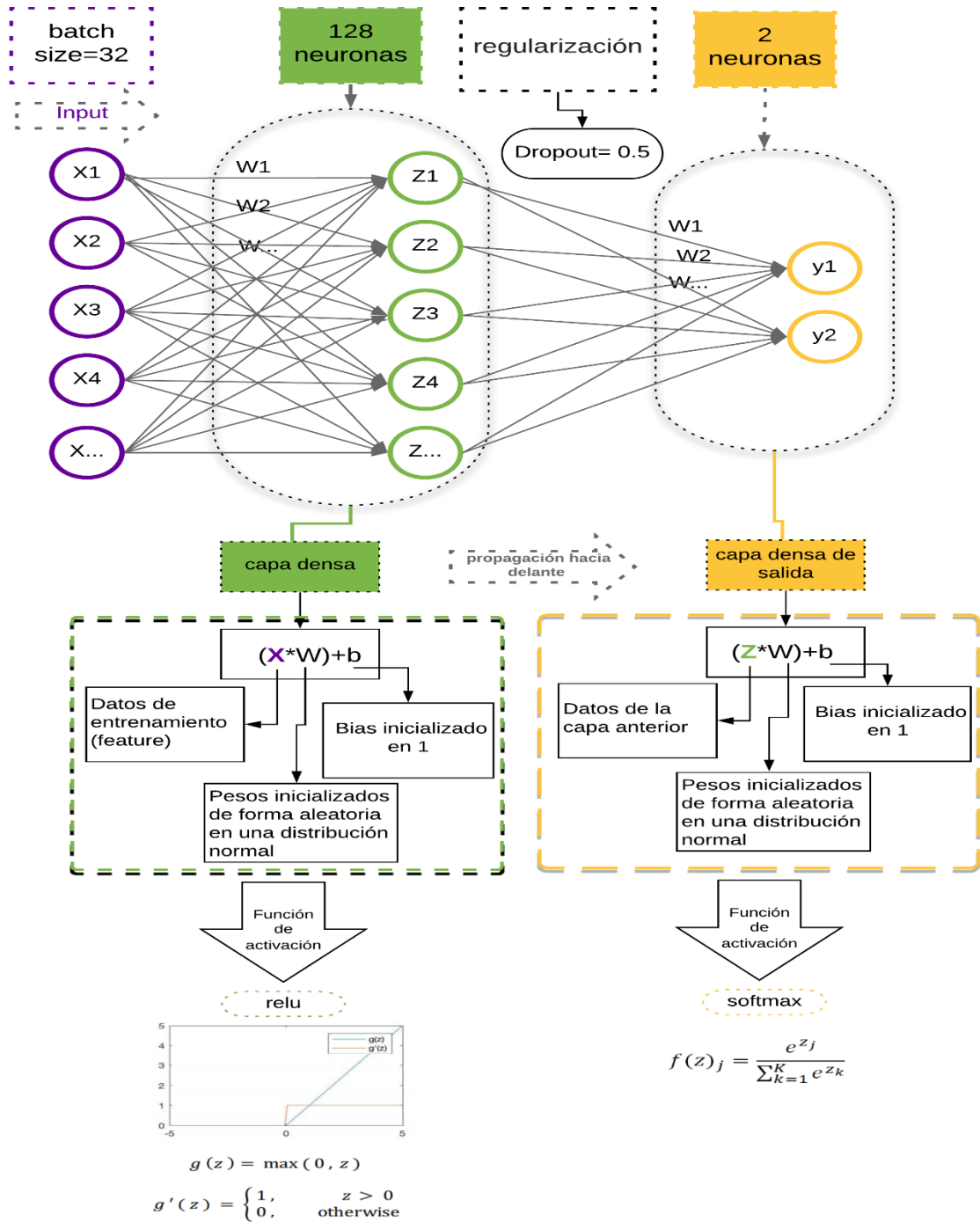


Ilustración 42 Construcción de la red neuronal profunda para la predicción de riesgos de MME

Como resultado se obtuvo una red neuronal densa compuesta de 2 capas densamente conectadas, es decir, que todas las neuronas de cada capa están conectadas con todas las neuronas de la capa siguiente, a esta primera capa densa (*dense*) se le asigna 128 neuronas, un *kernel_initializer* normal que define la forma de establecer los pesos, una función de activación *Relu* (ya que arroja mejores resultados que la sigmoide) y un *Bias_initializer ones* para generar tensores inicializados en 1.

Se tuvo que aplicar técnicas de filtrado de variables como la RFE y sus parámetros, se obtuvo un nuevo grupo de columnas de las cuales se seleccionarán las 20 más relevantes para el entrenamiento del modelo y por medio del filtrado *ExtraTreesClassifier()* se implementa un metaestimador que ajusta varios árboles de decisión aleatorios en varias submuestras del conjunto de datos, utilizando el promedio para mejorar la precisión predictiva y controlar el sobreajuste (*Overfitting*). También se tuvo en cuenta la técnica llamada *smote* para poder balancear los datos de la clase '0' y '1' ya que se encontraban desbalanceados.

Como resultado se obtuvo una cantidad de 31.930 registros de los cuales se dividen en 2 partes, 15.965 para la cantidad de '0' y 15.965 para la cantidad de '1' obteniendo un conjunto de datos para entrenamiento balanceados.

Los datos de pruebas también tuvieron que ser balanceados ya que para la cantidad de '0' tenían 6.910 valores y la cantidad de '1' 960 valores.

Como resultado se obtuvieron 31.930 datos para entrenamiento y 13.820 para prueba con 20 columnas y 2 clases de salidas '0 no MME' y '1 si MME'

Al estar los datos filtrados, balanceado y divididos y regularizados en los conjuntos de datos de entrenamiento y prueba, procede a realiza el entrenamiento del modelo obteniendo los siguientes resultados.

Para la compilación del modelo fue necesario probar diferentes optimizadores, el que más se ajustó al modelo fue el optimizador *adam* encargado de actualizar los pesos de red de forma iterativa en función de los datos de entrenamiento. La función de costo seleccionada fue la *categorical_crossentropy* la cual mide el rendimiento de un modelo de clasificación cuyo resultado es un valor de probabilidad entre 0 y 1. La métrica seleccionada para monitorear los pasos de entrenamiento y prueba es *accuracy* (exactitud), la fracción de los datos que son correctamente clasificados. El *learning rate* (tasa de aprendizaje) es el método que regula la velocidad para encontrar el mínimo local de error.

Para el entrenamiento del modelo se debieron ajustar los parámetros como el *batch_size* con un valor de 32 ya que van a hacer 32 muestras que se propagarán a través de la red neuronal.

Así mismo *validation_data* ayuda a seleccionar los datos sobre los cuales va a evaluar la pérdida y cualquier métrica del modelo al final de cada época.

Por otra parte, barajar (*shuffle=True*) baraja los datos de entrenamiento antes de cada época y El número de épocas (*epochs*) son 800 épocas para entrenar el modelo, realiza una iteración sobre todo *X_train* y *y_train* de datos proporcionados

Obteniendo como resultado

Resultados del con el bloque de entrenamiento			
Perdida	Exactitud	Validación de perdida	Validación de exactitud
10.77%	0.64%	0.68%	0.64%

Tabla 18 Resultado del bloque de entrenamiento del modelo

Los resultados no son los esperados ya que se espera que el margen de sensibilidad sea superior al 90% y una precisión por encima del 90%. En este modelo la exactitud obtenida fue del 64% y la pérdida de 10.77%.

Métricas para la evaluación del modelo

Los resultados arrojados por la matriz de confusión, cuentan con un total de VP= 4.461 (Verdaderos positivos), es significa que clasifico 4.610 que realmente tuvieron MME. VN = 4.201, lo cual indica que a 4.201 pacientes clasificó como sano o sin riesgo de MME, por otro lado FN = 2.300 se traduce a las clasificaciones incorrectas de los pacientes que llegaron a tener MME. Finalmente los falsos positivos FP = 2.709 son los pacientes a los que se les clasifico con riesgo para MME pero que en realidad no lo tenían.

El resultado de cada una de las métricas obtenidas mediante la matriz de confusión, se escogen para la medición del modelo dado que representan las predicciones positivas. Lo anterior se justifica tomando como premisa que para este proyecto es más importante acertar en las predicciones positivas que en las negativas. Por lo anterior las métricas que se usarán para este proyecto son: la exactitud, la precisión, la sensibilidad, la medida F1 y la prevalencia.

Métricas	Resultado
Exactitud	0.64%
Precisión	0.67%
Sensibilidad	0.63%
Medida F1	0.65%
Prevalencia	0.53%
Error	0.36%
AUC sensibilidad	0.64%
AUC Especificidad	0.28%

Tabla 19 Resultados de las métricas

Lo que se busca es un margen de sensibilidad sea superior al 90% con una precisión por encima del 90%. En este modelo la sensibilidad obtenida fue del 63% y la

precisión 64%. Lo cual indica que el modelo debe seguir entrenándose para poder ser usado por el sector salud.

CONCLUSIONES Y RECOMENDACIONES

Como resultado de este proyecto de investigación se realizó la construcción de una red neuronal profunda para la predicción de riesgos de Morbilidad Materna Extrema, esta red neuronal consta de 2 capas densas *dense* una capa con 128 neuronas y la otra capa con 2 neuronas. La primera capa de entrada se inicializa con un `kernel_initializer normal` el cual define la forma de establecer los pesos normal o aleatoriamente, así mismo la función de activación *Relu*, ya que se ajusta mejor al modelo que otras funciones de activación como la función sigmoide y un `Bias_initializer ones` para generar tensores inicializados en 1, finalizando así el ajuste de los parámetros del modelo *Sequential()* de la librería Keras.

Dando así cumplimiento al objetivo general planteado el cual era construir una red neuronal profunda para la predicción de riesgos de Morbilidad Materna Extrema.

Para cumplir con los objetivos específicos fue necesario seguir las fases planteadas por la metodología TDSP (Team Data Science Process) de Microsoft de las cuales se dividen en 5 fases y solo se llevarán a cabo 3 para este proyecto:

- Conocimiento del negocio,
- Adquisición y comprensión de los datos
- Modelado.

Las fases del conocimiento del negocio y la adquisición y comprensión de los datos abarcan el primer objetivo específico. Se comenzó con una entrevista con el ginecólogo de la clínica de maternidad Rafael Calvo fue de gran importancia ya que también se analizó el protocolo de morbilidad materna extrema, se analizó el procedimiento de una paciente en la etapa prenatal y se definieron los criterios de inclusión maternos sugeridos por el ginecólogo.

La obtención de los datos se realizó con la aprobación de la entidad de salud Clínica Maternidad Rafael Calvo, a fin de generar el set de datos para el entrenamiento de

la red neuronal. Al conjunto se le realiza un análisis de variables de riesgo para definir las variables a utilizar para entrada y salida del algoritmo, ya definido las variables dependientes e independientes, se pasan los datos a un software llamado SQL Server 2008 para realizar una limpieza de aquellos datos con ruido, es decir, campos vacíos, nulos, en fin, datos inapropiados para el conjunto de datos.

Por último, estos datos pasan a Excel, allí se hace la transformación para que queden separados por comas y se obtiene al final un archivo .CSV listo para cargarlo al algoritmo. Dando así por cumplido el primero objetivo específico la construcción del conjunto de datos.

Dando continuidad a la metodología planteada en este objetivo específico se aplica la fase de modelado ya que se hizo el entrenamiento del algoritmo para lo cual fue necesario aplicar técnicas de filtrado de variables como la técnica RFE y ExtraTreesClassifier, estas técnicas fueron aplicadas para conocer cuáles eran las variables más significativas del conjunto de datos, lo cual arrojó 30 columnas más significativas y por medio de ExtraTreesClassifier se seleccionaron las 20 más relevante para la técnica de filtrado. Así mismo fue necesario normalizar los datos para transformar los valores de los datos a un rango determinado de entre 0 a 1 de forma que el escalador organice por medio de la función *fit.transform()* transforma las columnas arrojadas por el algoritmo de filtrado *RFE* en un array bidimensional. con los datos filtrados y normalizados se realiza la división de los datos o *Split* el conjunto de datos se dividió en dos partes una para entrenamiento con el 70% y una parte para prueba del 30% con un total de 18.361 datos y 20 columnas de entrada y 1 de salida para el entrenamiento del modelo.

Por otra parte, los datos desbalanceados afectan al modelo en su proceso de generalización de la información por lo cual perjudica a las clases minoritarias. Para combatir este desbalance en los datos se utilizó la técnica smote de sobremuestreo de la minoría de clases, obteniendo así una igualdad en las clases '0' con 6.910 registros y '1' con 6.910 registros obteniendo unos datos para entrenamiento balanceados este mismo proceso se aplicó a los datos de prueba. Como resultado

se obtuvieron 31.930 datos para entrenamiento y 13.820 para prueba con 20 columnas y 2 clases de salidas '0 no MME' y '1 si MME'. con los datos balanceado es hora de construir el modelo.

Para la compilación del se utilizó la función *model.compile()*, con el optimizador *adam* encargado de actualizar los pesos de red de forma iterativa en función de los datos de entrenamiento. La función de costo seleccionada fue la *categorical_crossentropy*, la métrica seleccionada para monitorear los pasos de entrenamiento y prueba es *accuracy* (exactitud), la fracción de los datos que son correctamente clasificados y un *learning rate*.

En la etapa del entrenamiento del modelo se deben ajustar parámetros como él *batch_size*, *validation_data*, (*shuffle=True*), (*epochs*). Concluyendo así con el cumplimiento del segundo objetivo específico.

Siguiendo las fases de la metodología el tercer objetivo también es abarcado por la fase de modelado. En este objetivo se aplican métricas para poder evaluar el comportamiento del modelo en este caso se utilizó la exactitud y la perdida obteniendo 0.64% en exactitud y 0.69% en perdida durante el entrenamiento, al compararlo con el otro conjunto de datos destinado a pruebas se obtuvo una exactitud de 0.64% y una pérdida de 0.69%.

El resultado de cada una de las métricas obtenidas mediante la matriz de confusión, se escogen para la medición del modelo dado que representan las predicciones positivas. Lo anterior se justifica tomando como premisa que para este proyecto es más importante acertar en las predicciones positivas que en las negativas. Por lo anterior las métricas que se usarán para este proyecto son: la exactitud 0.64%, la precisión 0.67%, la sensibilidad 0.63%, la medida F1 0.65% y la prevalencia 0.53%. Así se concluye el tercer objetivo específico validar el modelo construido usando el 30% del conjunto restante para realizar los ajustes necesarios de los parámetros del modelo en términos del error.

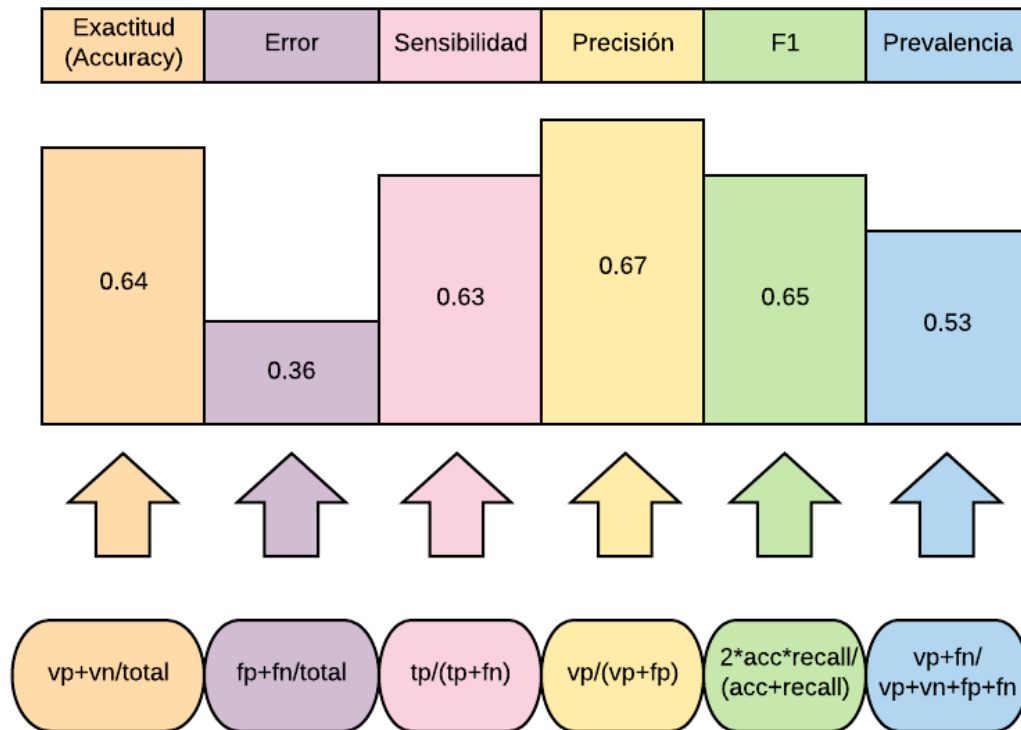


Ilustración 43 Resultado de las métricas

Se recomienda para futuros proyectos o la continuación de este aumentar el nivel de predicción y sensibilidad del modelo para mejores resultados, puesto que, los resultados obtenidos no cumplieron con los requisitos establecidos por el médico ginecólogo vinculado al proyecto, esto a causa del conjunto de datos entregado por la clínica por tal motivo se recomienda mejorar la calidad de los datos ya que en la mayoría de los casos no se encuentran debidamente estructurados y unificados lo que dificulta su análisis y la pérdida de datos para la investigación puesto que los médicos no realizan la digitación de los datos de manera adecuada relleno los campos con información irrelevante dificultando el manejo de los mismos por eso es necesario que se formateen los datos y la clínica verifique que se está cumpliendo con las normas al momento de diligenciar los formatos de registros clínicos. Otra sugerencia a tener en cuenta es ampliar el conjunto de datos a otras

entidades de salud de primer nivel y/o EPS, ya que en estas se prestan la primera atención lo que permite aumentar el volumen y evitar el desbalanceo de los datos puesto que en este proyecto se trabajó solamente con los datos proporcionados por la entidad de salud Clínica de Maternidad Rafael Calvo contando con un volumen de datos que limitó el resultado final de la investigación ya que requiere un número mayor de pacientes positivos y por último aplicar técnicas de aprendizaje profundo y de minería de datos que permita la construcción de un conjunto de datos más robusto y con un margen de error mínimo.

BIBLIOGRAFÍA

- [1] N. S. N. M. L. O. M. F. E. P. A. O. E. P. G. y. H. Q. B. Díaz, Morbilidad Materna Extrema, 2017.
- [2] C. A. M. Martínez, Caracterización de la Morbilidad Materna Extrema en el instituto materno infantil - hospital la victoria, Bogotá, 2010.
- [3] N. Muñoz, Informe Quincenal Epidemiológico Nacional. IQUEN, vol. 19, nº 17, p. 16., Bogotá, 2014.
- [4] A. I. H. D. Cancer., Ossola, A, Popular Science., 2016.
- [5] S. Mason, Microscope uses artificial intelligence to find cancer cells more efficiently., UCLA - Newsroom, 2016.
- [6] elimparcial, La muerte materna tiene un rostro de pobreza., Coahuila, 2017.
- [7] PROESAH, LUZi: Inteligencia Artificial contra la mortalidad materna e infantil, Guerrero: thestandardcio, 2018.
- [8] E. A. Rodríguez, *ANÁLISIS DEL COMPORTAMIENTO DE UNA RED NEURONAL PARA PREDICCIÓN DE MORBILIDAD MATERNA EXTREMA EN LA ETAPA DEL POSTPARTO, EN LA CLÍNICA DE MATERNIDAD RAFAEL CALVO*, cartagena, 2019.
- [9] E. Arrieta, J. C. Martínez y W. Caicedo, «early prediction of severe maternal morbidity using machine learning techniques,» *Ibero-American Conference on Artificial IntelligenceSpringer*, vol. 10022, pp. 259-270, 2016.
- [10] C. Radio, Luchando para disminuir cifras de mortalidad materna., Bolívar: https://caracol.com.co/emisora/2017/05/27/cartagena/1495838646_155347.html, 2017.
- [11] Google, Obtenido de Aprendizaje Automatico:., https://developers.google.com/machine-learning/crash-course/glossary?hl=es-419#sigmoid_function, 2018.
- [12] MinTic, Agencia Nacional del Espectro (ANE), Ministerio de Tecnologías de la Información y Las Comunicaciones, 2019.
- [13] P. (. Larranaga, Departamento de Ciencias de la Computación e Inteligencia Artificial., vasco: Universidad del País Vasco, 2012.
- [14] M. Azure, Obtenido de El Ciclo de vida del proceso de ciencia de datos en equipo, Microsoft Azure, 2017.
- [15] Microsoft, Metodología Time Data Science, Microsoft, 2018.

[16] I. N. D. V. E. S. P. (INS), Protocolo de morbilidad materna extrema, Colombia: <https://www.ins.gov.co>.

ANEXOS

CRONOGRAMA

	ACTIVIDAD	JULIO	AGOSTO				SEPTIEMBRE				OCTUBRE				
		1	2	3	4	5	6	7	8	9	10	11	12	13	
OBJETIVO 1	Entendimiento del negocio														
	Adquisición de los Datos														
	Comprensión de los Datos														
	Creación del Dataset														
OBJETIVO 2	Preparar entorno para entrenamiento														
	Conceptos e investigación de red neuronal														
	Construir algoritmo para minar los datos														
	Elaborar algoritmo con la red neuronal														
	Entrenamiento del modelo														
OBJETIVO 3	Validación del modelo														
	Analizar los resultados aplicando métricas al modelo														
	Optimizar parámetros														

ENTREGAS	Redacción de artículo Científico																
	Informe Técnico del Software																
	Descripción del Software																
	Monografías																
	Presentación Diapositivas																
	Código Fuente																