

Evaluación del riesgo de defunción fetal en embarazo de mujeres guatemaltecas por medio de modelos de aprendizaje automático

Diego Franco, fra20240@uvg.edu.gt / Daniel González, gon20293@uvg.edu.gt / Angel Carrera car20593@uvg.edu.gt
Departamento de Ciencias de la Computación, Facultad de Ingeniería, Universidad del Valle

Resumen: En este artículo, se analizaron los datos de registros de nacimientos y defunciones fetales de mujeres desde el 2010 al 2019 registrados en el Instituto Nacional de Estadística (INE) de Guatemala. Los datos constan de información sobre el departamento y el municipio donde se registró el nacimiento o la defunción, además de la edad de la madre, su escolaridad, su estado civil y su forma de atención médica, esto incluye, pero no se limita a: información del lugar de nacimiento, por ejemplo, una casa o un hospital; la atención que recibió durante el parto, por ejemplo, una enfermera o una comadrona. Adicionalmente se utilizaron modelos de predicción para poder saber si un embarazo terminaría en una defunción fetal o no. Los algoritmos de clasificación lograron generalizar bastante los datos logrando tener en promedio una precisión de 0.86 y el mejor modelo obtuvo una de 0.94. Estos modelos se evaluaron para poder asegurar que pudieran generalizar a datos nuevos, encontrando dos efectivos. De este modo se pudo encontrar una herramienta que busca ser una asistencia y ayuda para las mujeres, para mitigar el riesgo y actuar ante una emergencia durante su embarazo. También se evidenció la falta de estudios relacionados con las defunciones fetales en el país y la necesidad de estos para

poder desarrollar herramientas más avanzadas y precisas que mejoren la asistencia médica durante embarazos.

Palabras clave: Predicción de defunciones fetales, Factores de Embarazos Fallidos, Algoritmos de Aprendizaje Automático.

Evaluation of the risk of fetal death in pregnant Guatemalan women using machine learning models

Abstract: In this article, data from birth and fetal death registries of women from 2010 to 2019 registered in the National Institute of Statistics (INE) of Guatemala were analyzed. The data consists of information on the department and municipality where the birth or death was registered, in addition to the mother's age, her schooling, marital status, and her form of medical care, this includes, but is not limited to: information on the place of birth, e.g., a home or a hospital; the care she received during delivery, e.g., a nurse or a midwife. Additionally, prediction models were used to be able to tell whether a pregnancy would end in a fetal death or not. The classification algorithms were able to generalize the data quite well with an average accuracy of 0.86 and the best model with an accuracy of 0.94. These models were evaluated to ensure that they could generalize

to new data, and two were found to be effective. Thus, it was possible to find a tool that seeks to assist and help women to mitigate the risk and act in the event of an emergency during pregnancy. The lack of studies related to fetal deaths in the country and the need to develop more advanced and accurate tools to improve medical assistance during pregnancies was also evidenced.

Key words: Prediction of fetal deaths, Factors of Failed Pregnancies, Machine Learning Algorithms.

Introducción

Una defunción fetal según Mattingly (2022) es la pérdida de un feto en cualquier etapa prenatal. Se produce cuando luego de la completa expulsión o extracción del feto de la madre, este no respira ni muestra otra evidencia de vida, como pulso o movimientos voluntarios musculares.

Esta situación tiene diversas implicaciones. La primera es la posibilidad de muerte de la madre durante o previa al parto. También, los padres que sufren de una defunción fetal tienen índices superiores de depresión y ansiedad. Las madres tienden a aislarse y sufrir síntomas psicológicos que pueden prolongarse ilimitadamente, como menciona OMS (2020).

Según los datos provistos por el INE, en Guatemala se estima una tasa de 8.13 defunciones fetales por cada 1000 embarazos. En contraste en países desarrollados se producen 2.9 defunciones por cada 1000

embarazos, según Kestler y Ambrosio (2022). A pesar de esto, no se tienen suficientes fuentes de datos que permitan identificar claramente las causas de estas defunciones. Además, se tiene poco acceso a centros de salud, especialmente en el área rural, lo que hace que las defunciones fetales sean una situación alarmante en el país por las consecuencias planteadas y por la falta de atención médica de calidad.

Como evidencian Khatibi *et al.* 2021, se pueden crear modelos que incluyen árboles de decisión, bosques aleatorios y SVM que permitan realizar predicciones bastante acertadas para identificar posibles defunciones fetales. Sin embargo, se tienen conjuntos de datos que incluyen información de salud de los fetos, como su altura o peso. Por la misma razón de la falta de datos en Guatemala, no se tiene información tan detallada sobre las defunciones. Así mismo, no se ha desarrollado una herramienta en el país que permita asistir al área de salud para identificar un riesgo potencial de defunción fetal en embarazos.

Por lo tanto, el principal objetivo de esta investigación es utilizar modelos de aprendizaje automático para predecir si un embarazo será exitoso o no, usando principalmente variables demográficas relacionadas con la madre. También se busca mostrar las variables que más influyen en las defunciones fetales y patrones temporales en estas. Finalmente se busca encontrar la relación entre la edad de la madre, su educación y el resultado de un embarazo.

Marco Teórico

Situación problemática

Según estadísticas de OMS (2020), anualmente se producen cerca de dos millones de defunciones fetales en el mundo. Un 84% de estas se producen en países en vías de desarrollo. Se indica que la mayor cantidad de muertes son producidas por condiciones de salud precarias, específicamente en la falta de atención médica durante la etapa prenatal. Además, cerca del 40% de las muertes son producidas durante el parto, las cuales podrían prevenirse si se tuviera suficiente información de la etapa de gestación previa y personal adecuado para atender a la mujer embarazada.

En Guatemala también se da esta problemática. La atención médica es importante para la salud del feto, sin embargo, esta es precaria. Un estudio por Kestler y Ambrosio (2022) evidenció que la mayor cantidad de defunciones fetales se da en ámbitos no hospitalarios, con el 81%. Estos partos son atendidos por comadronas o familiares que pueden no tener la preparación necesaria. Algunos de los factores que influyen en la asistencia recibida por las madres pueden ser la edad, la escolaridad de la madre y el acceso a instalaciones adecuadas.

Por otra parte, según el MSPAS (2017) existe una brecha en la cantidad de emergencias atendidas en el área urbana y rural. 83% de los partos atendidos en el área urbana son atendidos en centros adecuados de salud,

mientras que en el área rural solo 55%. Desde luego, Guatemala es el departamento que más partos atiende en centros de salud, seguido por Escuintla y Santa Rosa.

Según los registros del INE, en Guatemala la falta de atención médica es un problema bastante común provocando alrededor de 2800 abortos anuales. La mayoría de los registros están en el departamento de Guatemala, porque se busca una mejor atención médica, a pesar de esto, los centros de salud no están lo suficientemente equipados para atender las emergencias de embarazos, conllevando a defunciones que se pudieron prevenir.

Por ende es muy importante centrarse en que la atención que se le brinda a la madre antes, durante y después del parto sea la adecuada. Esto incluye el uso de las técnicas adecuadas, que los centros de salud estén equipados con las herramientas necesarias para que la supervisión sea la óptima.

Conceptos de algoritmos utilizados

Para la investigación se utilizaron diversos algoritmos de aprendizaje automático de clasificación, los cuales tienen ciertas características que se mencionan a continuación.

Support Vector Machine (SVM), es un algoritmo de aprendizaje supervisado que se utiliza en muchos problemas de clasificación y regresión. El objetivo del algoritmo SVM es encontrar un hiperplano que separe de la mejor

forma posible dos clases diferentes de puntos de datos.

El funcionamiento de este algoritmo se centra en correlacionar los datos en un espacio de grandes dimensiones con el fin de poder categorizarlos. Para poder realizar esta separación se debe de detectar un separador con el cual se hacen las transformaciones necesarias a los datos por medio de la función matemática kernel, y así extraer los datos como un hiperplano. Así lo define IBM (2021)

En la investigación también se usó el algoritmo de *XGBoost*. Este modelo se basa en el principio de boosting que consiste principalmente en generar múltiples modelos de predicción que no son muy buenos. Como indica Mendoza (2020) se usa una estrategia greedy, que consiste en seleccionar el modelo que dé mejores resultados. Luego se emplea un algoritmo de optimización conocido como gradiente descendiente que minimice el error producido por los modelos, modificando sus parámetros. El proceso se termina cuando todos los modelos tienen desempeños muy similares, lo que indica que se encontró el mejor.

Según Chen y Guestrin (2016), *XGBoost* está optimizado para poder usar computación paralela, de forma que se puedan procesar datos simultáneamente. Esto mejora considerablemente su velocidad. Además, usa un método de memoria caché para almacenar estadísticas relacionadas al conjunto de datos,

de forma que no se tengan que computar si ya fueron obtenidas previamente.

Otro algoritmo utilizado dentro de la investigación fue el *Random Forest*. Este algoritmo utiliza múltiples árboles de decisión sobre el conjunto de datos utilizado, se combinan todos los resultados de los árboles para obtener un modelo mucho más robusto. Como menciona Breiman (2001), para utilizar este algoritmo es necesario pasar por un proceso de dos etapas. La primera es crear un número de árboles de decisión considerable según el conjunto de datos. Luego cada uno de los árboles debe crecer hasta su máxima extensión, para luego combinar las soluciones.

Además se utilizó el algoritmo de regresión logística, algoritmo de clasificación que predice la probabilidad de una variable dependiente categórica. Como indica Gonzalez (2019) este algoritmo es uno de los más simples y utilizados porque utiliza como su nombre indica una función logística con la cual se crea una curva para hacer la predicción de a qué grupo pertenecen los datos. Si la curva tiende al infinito positivo el valor se convertirá en 1. Por el contrario si la curva tiende al negativo el valor se convertirá en 0.

Finalmente, se empleó una red neuronal artificial prealimentada de una sola capa. Como menciona Sharma (2022) una red neuronal es un modelo que se basa en capas que contienen múltiples neuronas. En este caso, la arquitectura prealimentada se basa en una sola capa que puede tener una cantidad

variable de neuronas. Cada neurona tiene asociada una función de pesos y un *bias*. Esto se une con una función de activación, la cual busca simular el funcionamiento del cerebro humano para transmitir impulsos eléctricos. La función toma como input el resultado de aplicar la función de pesos a la entrada del modelo y devuelve cierto valor. Este es usado para estimar la clasificación de la entrada del modelo.

Materiales y métodos

Los datos se conformaron por 10 años de registros, que van de 2010 a 2019, de embarazos y defunciones fetales, por separado. Estos registros fueron proporcionados por el sitio de estadísticas vitales del INE de Guatemala. Los datos estaban separados por año, por lo que se debió de realizar una unión de estos en un solo conjunto. En total, este conjunto de datos original tenía 3,842,562 datos, de los cuales el 99.19% de los datos eran de embarazos exitosos.

Debido a este desbalance de datos se realizó una modificación al conjunto de datos. Esto consistió en tomar una muestra aleatoria de 31,228 datos de los registros de embarazos exitosos. Esta cantidad de datos fue la misma que se tenía de datos de defunciones fetales. De esta forma, se logró balancear la cantidad de registros de cada clase de registro. Hubo ciertas diferencias entre los conjuntos de variables para los registros de cada año. Por lo tanto, hubo que encontrar las variables comunes de todos los datos. Inicialmente se identificaron 21 variables que se encontraban

en todos los registros. Luego se procedió a unir todos los registros en un solo conjunto de datos.

Las variables del conjunto de datos incluyen información sobre el departamento y municipio de registro del embarazo o defunción, la edad de la madre, su escolaridad y estado civil. También se tuvo información de el lugar donde ocurrió, por ejemplo una casa o un hospital. Finalmente se tuvieron datos relacionados a la cantidad de hijos de la madre y si están vivos o han fallecido. De las variables hubo cinco numéricas y dieciséis cualitativas.

Tratamiento de datos

Luego de haber delimitado cuál iba a ser el dataset completo, tuvimos que realizar diferentes transformaciones para poder garantizar la calidad y adecuación de los datos. Como primer tratamiento realizamos un análisis de valores faltantes, se verificaron cuales eran las variables que no estaban completas y decidimos eliminar las observaciones que correspondían a estos datos. Al realizar esto nos aseguramos de tratar únicamente con datos completos.

La segunda evaluación que realizamos fue identificar las variables que iban a ser realmente significativas para la investigación. Para poder hacer esto realizamos un análisis exploratorio de las cada una de las variables que quedaron en el dataset, se fue evaluando una por una para saber si la información que ésta nos proporcionaba era relevante para

completar el objetivo de esta investigación. Por ejemplo, en el dataset original se tenían las variables (TOHIVI) Total de hijos vivos, (TOHITE) Total de hijos tenidos, (TOHINM) Total de hijos nacidos muertos. Pero al verificar la correlación entre estas variables (Fig. 1) nos dimos cuenta que el tener la variable de TOHITE realmente no iba a ser significativo porque era casi lo mismo que tener las otras dos variables. Por lo que decidimos dejar únicamente las variables que consideramos más significativas y que nos ayudarían a completar la investigación en cuestión.

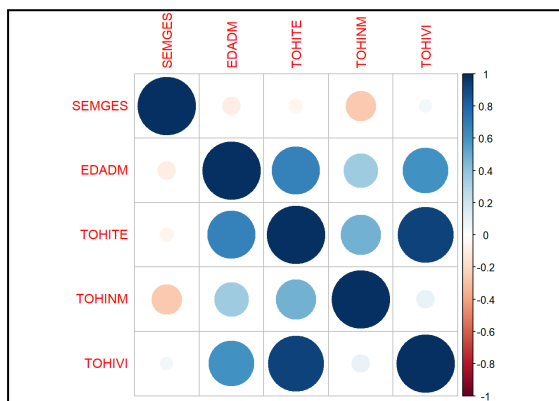


Figura 1. Correlación de las variables del dataset.

La alta cantidad de registros en el departamento de Guatemala y sus municipios, causaron un sesgo en los datos. Debido a que esto podía afectar las predicciones de los datos, se decidió que estas variables no iban a ser representativas en el resultado de la investigación y no fueron utilizadas durante la aplicación de modelos.

Por otra parte, al crear el conjunto de datos se incluyó la variable de semanas de gestación

(SEMGES), sin embargo, al crear el dataset se asignó un valor de 40 a todos los embarazos exitosos. Esto porque es el promedio de semanas de gestación para un nacimiento. A pesar de esto, se decidió eliminarla porque podía no ser representativa.

Luego de realizar la limpieza de datos, se evidenció cierto desbalance en las clases, teniendo aproximadamente un 67% de datos de nacimientos y un 33% de defunciones fetales. Sin embargo, se decidió no realizar un muestreo estratificado, debido a que la diferencia no es tan grande en los datos. Además, como se menciona más adelante, se usó F1 como métrica para penalizar esto.

Recursos de hardware y software

La investigación se realizó sobre una Alienware M15 R6 con un sistema operativo Windows 11 Home y un procesador Intel Core de 11th Generación i7-11800H y 32 GB de RAM. Se utilizó la versión de R 4.2.3 para el análisis de los datos y las siguientes librerías: nortest, versión 1.0-4; dplyr, versión 1.1.1; hopkins, versión 1.0; factoextra, versión 1.0.7; ggrepel, versión 0.9.2; cluster, versión 2.1.4; flexclust, versión 1.4-1; FeatureImpCluster, versión 0.1.5; stringr, versión 1.5.0; tidyr, versión 1.3.0; stats, versión 4.2.3; graphics, versión 4.2.3; NbClust, versión 3.0.1; mclust, versión 6.0.0; GGally, versión 2.1.2; corrplot, versión 0.92; caret, versión 6.0-93; ggplot2, versión 3.4.1; kableExtra, versión 1.3.4; e1071, versión 1.7-11; rpart, versión 4.1.19; rpart.plot, versión 3.1.1; naivebayes, versión 0.9.7; randomForest, versión 4.7-1.1;

RColorBrewer, versión 1.1-3; ca, versión 0.71.1; vcd, versión 1.4-11; xgboost, versión 1.7.3.1; mlr, versión 2.19.1; glmnet versión 4.1-7.

Selección de algoritmos

Teniendo los datos necesarios y el entorno completo, decidimos utilizar distintos algoritmos para poder crear la herramienta de predicciones.

El primer algoritmo usado fue Support Vector Machine. Este algoritmo fue elegido ya que tiene una gran capacidad para manejar relaciones no lineales por medio del uso de dimensiones superiores aplicadas sobre los datos y su resistencia al sobreajuste. Luego la red neuronal se seleccionó debido a su gran flexibilidad en la representación de características y su potencial para encontrar relaciones complejas entre los datos. Por otra parte, XGBoost se eligió principalmente por su potencia ya que es necesario que pueda procesar la cantidad de datos que se están manejando de manera rápida y eficiente. Además de su capacidad de reducción de sobreajuste y que es uno de los algoritmos más efectivos que han producido resultados del estado del arte. Luego, Random Forest se eligió, ya que es un algoritmo que maneja bien los datos faltantes y a pesar de haber hecho la intervención necesaria sobre los datos faltantes al tratarse de un tema de salud es muy común que los datos estén incompletos. Además, la unión de muchos modelos le permiten identificar relaciones complejas. Finalmente, la regresión logística, se seleccionó porque con

los datos que se tiene podría brindar no solo la clasificación final del embarazo sino también la probabilidad de que el embarazo sufra una emergencia.

Para evaluar la efectividad de los modelos, se aplicaron múltiples métricas y técnicas de análisis. Las métricas clave incluyen la precisión (*accuracy*) y la medida F1, mientras que las técnicas de análisis de datos comprendieron el uso de una matriz de confusión.

Los modelos se examinaron inicialmente en términos de la precisión, que es una métrica que proporciona una visión general de la capacidad del modelo para predecir resultados con exactitud. En este contexto es la competencia del modelo para clasificar e identificar correctamente las defunciones fetales. La precisión es una métrica que permite determinar la efectividad general del modelo al cuantificar el porcentaje de clasificaciones correctas. Sin embargo, la precisión por sí sola no proporciona una visión completa de la efectividad del modelo cuando los datos están desbalanceados.

Por lo tanto ya que dentro de nuestro conjunto de datos existe un desbalance, como anteriormente lo mencionamos, también se utilizó la medida de F1. Esta métrica se emplea para evaluar simultáneamente la precisión y la sensibilidad del modelo. En el caso del estudio presente, la medida de F1 se utilizó para determinar la precisión de las predicciones positivas del modelo. Esto es

fundamental para garantizar que el modelo no solo pueda identificar correctamente los casos de embarazos fallidos si no que también poder predecir efectivamente las predicciones positivas reales.

Resultados y discusión

Análisis Exploratorio

Dentro del análisis exploratorio que se realizó para determinar cuales variables eran más significativas también se obtuvieron los siguientes hallazgos

Se pudo observar que los registros de defunciones han ido disminuyendo a lo largo de los años. En el año 2010 se tenían más de 3000 registros de defunciones pero en el año 2019 solo se registraron cerca de 27 mil defunciones. Sin embargo, estos datos podrían no reflejar la cantidad de defunciones en el país, lo cual se sustenta por la falta de estudios y registros actuales.

También se pudo realizar una serie de tiempo para analizar cómo se dan las defunciones con respecto al tiempo. Igualmente, se evidenció una tendencia bajista en la cantidad de defunciones registradas. La parte más importante fue la estacionalidad que presentó el conjunto de datos. Se pudo observar un patrón que se repite aproximadamente cada año en las defunciones fetales (Fig. 2). Se puede observar un aumento importante durante los últimos meses del año.

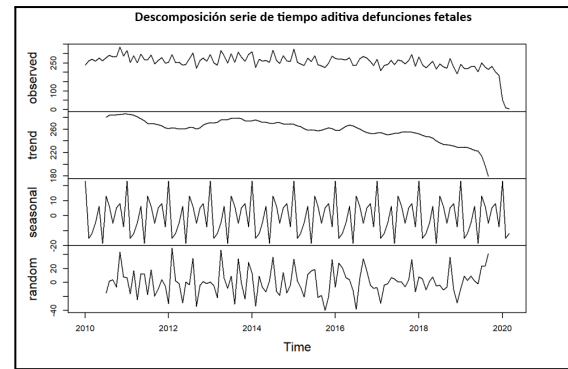


Figura 2. Serie de tiempo de defunciones fetales.

Se logró determinar que aproximadamente el 70% de las defunciones fetales son tratadas en centros médicos. El otro 30% está distribuido en atención dada por comadronas la cual es mucho menos adecuada cuando se trata de un embarazo que tiene riesgo y por último tenemos los casos en donde la difusión no recibió nada de atención. Esto nos permite identificar que la atención dada a los casos de embarazos peligrosos, a pesar de ser dada por médicos es inadecuada haciendo que la defunción no se pueda evitar.

Por otra parte, se hizo una análisis de la relación entre la escolaridad de la madre y el resultado del embarazo. Inicialmente se hizo una prueba chi-cuadrado que indicó que sí había relación. Sin embargo, al realizar una prueba de correspondencia con la medida V de Cramer, esta resultó ser de 0.097, un valor demasiado bajo. Por lo tanto, no es una variable que influya mucho en el resultado del embarazo. Igualmente, se buscó encontrar la relación entre la edad de la madre y el resultado del embarazo, pero la medida V arrojó un valor de 0.15, igualmente bajo, por

lo que no tiene mucha influencia. A pesar de esto, se evidenció que mientras más edad tiene una mujer, menos probabilidad existe de tener nacimientos y defunciones fetales.

Adicionalmente, se realizó un árbol de decisión simple en el cual se buscó encontrar las variables más relevantes al momento de predecir una defunción fetal. Se encontró que la principal es la cantidad de hijos nacidos muertos de una madre, seguida por la cantidad de hijos tenidos, la asistencia recibida y el tipo de parto.

También se determinó por medio del estadístico de Hopkins el cual obtuvo un valor de 1, que si era necesario hacer agrupamientos de datos. Tras hacer el método de la silueta y WSS, para determinar grupos en *clustering*, se determinó que la mejor cantidad de grupos en la que se debe de hacer la separación de los datos es 2. A pesar de ello a simple vista los grupos no se pueden distinguir con tanta facilidad (Fig. 3).

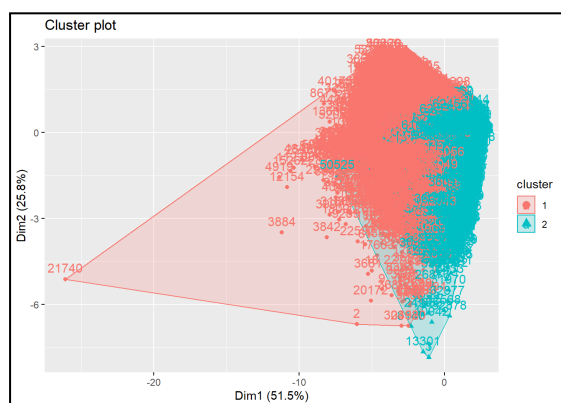


Figura 3. Agrupamiento de los datos.

Al hacer un mejor análisis sobre estos dos grupos se pudo determinar que el primer grupo son aquellas mujeres que tienen una edad relativamente alta en promedio 33 años, que su nivel educativo es prácticamente nulo, están casadas y tienen una familia numerosa con un promedio de casi 5 hijos en total. El segundo grupo son aquellas mujeres jóvenes en promedio 22 años, que en su gran mayoría alcanzan el nivel primario de educación, son solteras y tienen en promedio 2 hijos en total. Por lo que el grupo número uno se le denomina: “Madres de una edad avanzada, casadas, sin escolaridad, con una cantidad alta de hijos tenidos.”. Y al segundo grupo se le denomina como: “ Madres jóvenes, solteras, con educación primaria, con una cantidad relativamente baja de hijos.”.

Resultados de modelos predictivos

La tabla 1 muestra los resultados obtenidos en las principales métricas aplicadas para evaluar la clasificación de los algoritmos. Usamos un 80% de datos de entrenamiento y un 20% de prueba. Es posible observar que en cuanto a accuracy, la red neuronal y el random forest fueron los modelos que mejor se desempeñaron. En cuanto a sensibilidad, la regresión logística y la SVM fueron las que mejor pudieron identificar casos de verdaderos positivos. El bosque aleatorio fue el mejor para identificar verdaderos negativos y en F1.

El conjunto de datos era complejo, debido a que se tenían pocas variables que explícitamente pudieran tener relación con la variable respuesta. Sin embargo, la capacidad

de ajustarse a patrones complejos, tanto por parte del random forest como de la red neuronal, les permitieron ser buenos clasificadores. Además, estos dos modelos pudieron manejar adecuadamente el ligero desbalance de datos mencionado anteriormente, lo cual se puede observar en que tuvieron los mejores valores de F1.

La SVM tuvo valores buenos de accuracy y de F1. Fue el mejor modelo para detectar verdaderos positivos, lo cual se puede deber a que se encontró una buena separación por parte del hiperplano. Además, el uso del kernel lineal pudo ayudar a encontrar esta separación. También fue efectivo en el manejo del desbalance de datos, lo que evidencia su alto valor de F1.

Este mismo desbalance afectó a la regresión logística, que no tuvo una buena accuracy. Sin embargo, se puede observar que no fue un modelo malo para la clasificación porque su F1 no es tan distinto a los demás modelos. A pesar de esto, fue el peor modelo de todos. Se puede observar que tuvo una sensibilidad de 1 y una especificidad de 0. Esta diferencia drástica se puede deber al umbral de probabilidad usado, 0.5, para realizar la clasificación. Es muy probable que no haya sido el adecuado, lo que causó esta variación en la clasificación y lo que hizo en parte que fuera el modelo menos efectivo.

XGBoost fue un modelo regular al momento de realizar la clasificación de datos. Esto fue un resultado inesperado, porque generalmente

este modelo es muy apto para la complejidad de los problemas. Es posible que no se hayan tenido las características necesarias para que pudiera realizar predicciones adecuadas sobre el resultado de un embarazo. También pudo afectar el desequilibrio de la variable objetivo mencionada.

	Accuracy	Sensibilidad	Especificidad	F1
Red neuronal	0.9456	0.9896	0.926	0.9113
XGBoost	0.7924	0.836	0.7014	0.8449
Random Forest	0.9531	0.9687	0.9456	0.9315
Regresión logística regularizada	0.6764	1	0	0.807
SVM	0.9393	1	0.9103	0.9143

Tabla 1. Métricas de desempeño de modelos de aprendizaje automático.

Al saber cuál fue el rendimiento de los algoritmos utilizados es necesario discutir también sobre cuál de estos es realmente significativo para la investigación porque no es de utilidad que el modelo se haya adecuado únicamente a los datos con los que se está trabajando o en otras palabras que se haya sobre ajustado. Para poder saber si el modelo sufrió de algún sobre ajuste o desajuste se utilizó la curva de aprendizaje.

Para el algoritmo de support vector machine se obtuvo una curva de aprendizaje bastante irregular (Graf. 1). En esta se puede observar como la curva de entrenamiento y la de prueba se interceptan entre sí en diferentes ocasiones, lo que nos indica que el modelo se ha ajustado demasiado a los detalles y patrones de los datos de entrenamiento. Esto indica que el modelo está sobre ajustado.

Es posible que el hiperplano encontrado para la división de datos haya sido demasiado rígido que causó el sobre ajuste. También puede que los datos no hayan sido lo suficientemente significativos para poder realizar una clasificación adecuada de la variable. Sería importante buscar realizar una relajación de la división de datos por medio de tuneo de parámetros para mejorar el resultado.

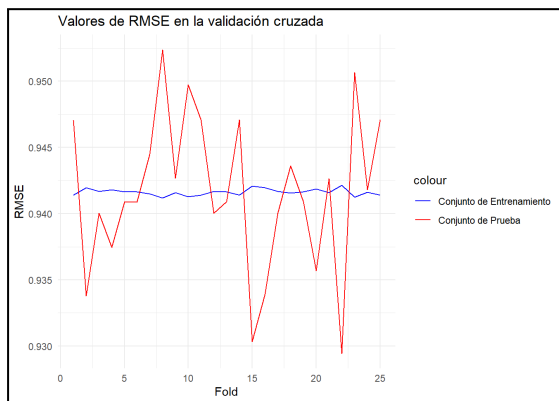


Gráfico 1. Curva de aprendizaje de SVM.

El siguiente modelo evaluado fue la red neuronal, del cual se obtuvo la siguiente curva de aprendizaje (Graf. 2) en donde claramente se muestra que las curvas de prueba y entrenamiento tienen un comportamiento muy similar. Aunque hay variaciones en el rendimiento a lo largo de las curvas, estas permanecen muy cerca una de la otra. En consecuencia, se puede concluir que las curvas convergen, lo que indica que el modelo no sufre de sobre ajuste.

Es importante mencionar que se trata de una red neuronal con una sola capa de 5 neuronas. Se realizó parameter tuning, para encontrar

que modelo se desempeñaba mejor y se obtuvo que se debían usar 10 neuronas. Sin embargo, con esta arquitectura se producía overfitting. Esto porque la complejidad del modelo aumentaba y se ajustaba demasiado a los datos de entrenamiento. Sin embargo, solo con las neuronas usadas, se produjo un buen modelo sin sobre ajuste.

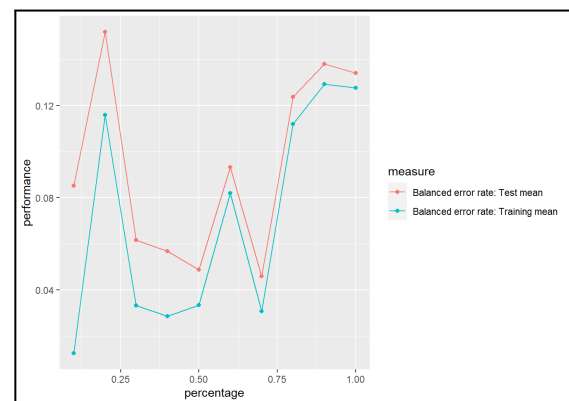


Gráfico 2. Curva de aprendizaje de red neuronal.

La siguiente curva de aprendizaje (Graf. 3) pertenece al modelo XGBoost. Al analizar la gráfica, se puede evidenciar que el modelo no sufre de sobreajuste debido a que, a pesar de que las curvas no sean idénticas, estas convergen. Esta convergencia indica que el modelo no está sobre ajustando a los datos de entrenamiento y es capaz de realizar una buena clasificación.

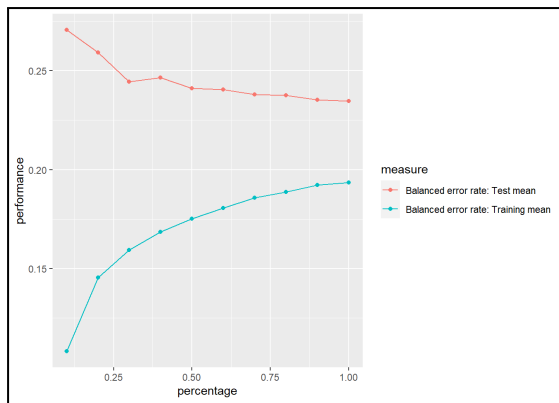


Gráfico 3. Curva de aprendizaje de XGBoost .

En cuanto a la curva del modelo random forest (Graf. 4), se puede apreciar que el error de entrenamiento continúa disminuyendo a medida que se incrementa la cantidad de datos utilizados para entrenar el modelo. Sin embargo, el error de validación se estabiliza y no muestra una mejora significativa. Esta observación sugiere que el modelo está memorizando los datos de entrenamiento y no logra generalizar de manera efectiva con nuevos datos. Esto nos indica que el modelo sufre de sobreajuste, implicando que su capacidad para generalizar y realizar predicciones precisas con nuevos datos se vea comprometida.

Este sobreajuste se puede deber a que el modelo es bueno para ajustarse a la complejidad de los datos, permitiéndole entender patrones del conjunto de entrenamiento. Sin embargo, esto mismo le reduce su capacidad de generalizar. En general este modelo no es adecuado para la solución del problema, debido a que se realizaron métodos para intentar reducir el sobreajuste, como aumentar la cantidad de árboles y

reducir la profundidad de cada uno y se obtuvieron los mismos resultados.

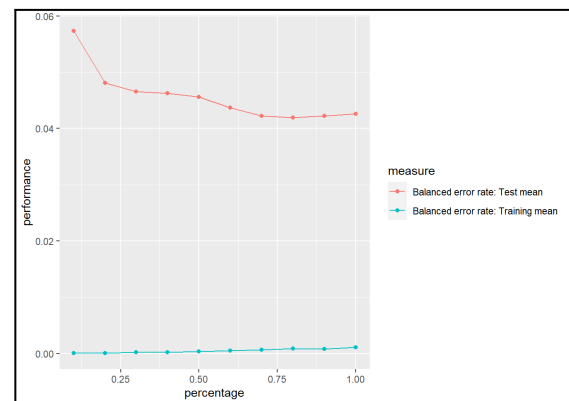


Gráfico 4. Curva de aprendizaje de random forest.

Por último se encuentra la curva de la regresión logística (Graf. 5). Como se puede observar las curvas son prácticamente idénticas pero a pesar de ellos se tiene una disminución significativa en ambas de forma repentina, lo que podría indicar la presencia de un problema de desajuste en este modelo.

Anteriormente en las métricas se evidenció que la especificidad del modelo era muy mala. Esto, unido con la evidencia de la gráfica de aprendizaje, evidencia el desajuste, debido a que no el modelo no está pudiendo aprender adecuadamente las relaciones entre datos. El modelo se usó con regularización, lo que buscó reducir el sobreajuste pero es probable que haya causado el efecto contrario.

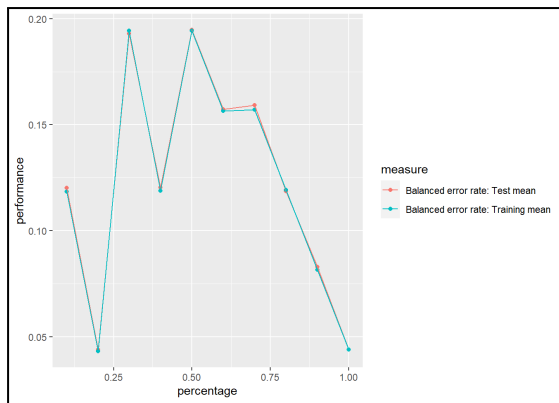


Gráfico 5. Curva de aprendizaje de regresión logística.

Después de analizar cada una de las métricas y tomando en cuenta únicamente aquellos modelos que proporcionaron un aporte significativo a la investigación, es decir los modelos que no sufrieron un sobre ajuste de los datos ni un desajuste. Se determinó que el mejor modelo es la Red Neuronal.

Conclusiones

La Red Neuronal demostró un rendimiento destacado en diversas métricas evaluadas, superando a los demás modelos considerados. Su capacidad para capturar patrones complejos y adaptarse a la estructura de los datos ha resultado en una gran clasificación de los mismos. Demostrando así ser la opción más sólida y prometedora en términos de rendimiento y aporte a la investigación. Es por ello que este algoritmo es la herramienta que se estaba buscando para poder determinar si una mujer tiene riesgo de perder al bebé en la etapa pre natal o en el parto.

Después de llevar a cabo el análisis exploratorio de los datos, se pudo establecer que dentro de las variables más significativas

para predecir el resultado de un embarazo esta la cantidad de hijos nacidos muertos, esta variable se posiciona en primer lugar debido a su fuerte impacto en la predicción por tener una correlación directa con la posibilidad de un desenlace adverso. El tipo de parto y la asistencia recibida también son factores importantes al momento de predecir el resultado de un embarazo. Esto enfatiza la necesidad de un buen sistema de salud que permita asistir a las madres.

En cuanto a datos más específicos, se pudo observar que la edad de la madre no desempeña un papel significativo tanto en la defunción fetal como en el nacimiento del bebé. Sin embargo, se pudo identificar una tendencia que sugiere que a medida que la edad de la mujer aumenta, se observa una disminución tanto en la cantidad de nacimientos como en las defunciones.

Por otro lado, se determinó que el nivel educativo de la madre tiene una correspondencia aún menor que la edad. Por lo tanto, se puede afirmar que el nivel educativo de la madre no influye significativamente en el resultado del embarazo.

En general, el proceso de clasificación de defunciones fetales es delicado y requiere de un análisis complejo de las variables que las pueden influenciar. A pesar de que el modelo de red neuronal brindó un resultado muy positivo, se sugiere para futuras investigaciones encontrar conjuntos de datos que cuenten con información más detallada del

estado del feto al momento de la defunción o del embarazo. De esta forma, se podrán evaluar factores que pueden influenciar más directamente la categorización y crear una herramienta mucho más precisa.

Bibliografía

- Aprende IA website:
<https://aprendeia.com/algorithmo-regresion-logistica-machine-learning-teoria/>
- Breiman, Leo. "Random Forests." *Machine Learning*, vol. 45, no. 1, 2001, pp. 5–32,
<https://doi.org/10.1023/a:1010933404324>.
- Chen, T., y Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 785–794.
<https://doi.org/10.1145/2939672.2939785>
- Gonzalez, L. (2019, Junio 28). Regresión Logística - Teoría - Aprende IA. Retrieved May 23, 2023, from
- IBM Documentation. (2021, August 17). Retrieved May 23, 2023, from <https://www.ibm.com/docs/es/spss-modeler/saas?topic=models-how-svm-works>
- Khatibi, T., Hanifi, E., Sepehri, M. M., & Allahqoli, L. (2021). Proposing a machine-learning based method to predict stillbirth before and during

delivery and ranking the features: nationwide retrospective cross-sectional study. *BMC Pregnancy and Childbirth*, 21(1).
<https://doi.org/10.1186/s12884-021-03658-z>

- Kestler, E., y Ambrosio, G. (2022). Nacido muerto en la comunidad y su relación con la morbilidad materna aguda y grave. *Revista Médica (Colegio de Médicos Y Cirujanos de Guatemala)*, 161(2), 105–112.
<https://doi.org/10.36109/rmg.v16i2.484MSPAS>. (2017). VI
- Mattingly, P. J. (2019, November 9). Evaluation of Fetal Death: Definition of Fetal Death, Frequency of Fetal Death, Diagnosis of Fetal Death. *Medscape.com*.
<https://emedicine.medscape.com/article/259165-overview>
- MSPAS. (2017). VI Encuesta Nacional de Salud Materno Infantil (ENSMI) 2014-2015 Informe Final.
<https://www.ine.gob.gt/sistema/uploads/2021/04/27/20210427195413QDinUvuRa9GjopyXaTuNMXc3gd6Jq1Q1.pdf>
- OMS. (2020, Octubre 8). Cada 16 segundos se produce una muerte fetal, según las primeras estimaciones conjuntas de las Naciones Unidas. *Www.who.int*.
<https://www.who.int/es/news/item/08-10-2020-one-stillbirth-occurs-every-16-seconds-according-to-first-ever-joint-un-estimates>

Sharma, P. (2022, March 7). Basic Introduction to Feed-Forward Network in Deep Learning. Analytics Vidhya.

<https://www.analyticsvidhya.com/blog/2022/03/basic-introduction-to-feed-forward-network-in-deep-learning/>

Vega, J. B. M. (2022, Enero 24). Tutorial: XGBoost en Python. Medium.
<https://medium.com/@jboscomendoza/tutorial-xgboost-en-python-53e48fc58f73>