

Minería de datos para la detección de factores de influencia en el test Apgar

Soledad Retamar, Anabella De Battista
Lautaro Ramos, Juan Pablo Nuñez
Depto. Ingeniería en Sistemas de Información
Fac. Reg. Conc. del Uruguay, Univ. Tecnológica Nacional
Entre Ríos, Argentina
{retamars, debattistaa, ramosl, nunezjp}
@frcu.utn.edu.ar

Francisco Savoy, Laura De Gracia
Fac. de Ciencias de la Salud
Univ. Nac. de Entre Ríos
Entre Ríos, Argentina
{savoyf, degracial}
@fcs.uner.edu.ar

Resumen

En este trabajo se aplicaron técnicas de minería de datos para clasificar los valores del test de Apgar en los nacimientos ocurridos en el hospital público Justo José de Urquiza entre los años 2009 y 2017. Se aplicaron técnicas de selección de atributos para reducir la dimensionalidad en la vista minable, se realizaron experimentos con tres algoritmos de balanceo de clases: Class Balancer, SMOTE y Spread Sub Sampled; y se aplicaron tres modelos de clasificación basados en árboles de decisión: J48, REP Tree y Random Tree. Se compararon los resultados mediante métricas como TPRate, F-Score y matriz de confusión obteniendo mejores resultados con la combinación de Class Balancer y Random Tree. Entre los principales factores resultantes de aplicar el modelo de clasificación se encuentran: la aplicación de oxitócicos durante el trabajo de parto, la posición del niño al nacer, el peso y la edad gestacional del recién nacido.

1. Introducción

Las afecciones originadas en el período perinatal y los defectos congénitos de los neonatos, continúan siendo la principal causa de muerte en Argentina en los menores de un año, representando más del 80% de las muertes reportadas [1], y conformando uno de los grandes problemas a los que se enfrentan las estrategias en salud pública.

El test Apgar fue creado con el objetivo de evaluar, por medio de una herramienta sencilla, la condición física del neonato inmediatamente después de haber

nacido en función de características obstétricas, disminución del dolor en la madre y los efectos de la resucitación [2]. Esta evaluación se repite nuevamente a los cinco minutos posteriores al nacimiento asignando al neonato un puntaje que varía en una escala de 1 a 10, siendo el 10 el puntaje óptimo de la escala. Se considera que si el recién nacido obtiene un valor del test de Apgar igual o superior a 7 se encuentra en buenas condiciones de salud, si el puntaje está por debajo de 7 en bajas condiciones y por debajo de 4 puntos en muy bajas condiciones. Los puntajes inferiores a 7 en el test de Apgar, especialmente a los cinco minutos de vida, indican un compromiso para la salud del recién nacido [3].

En este trabajo se analizan mediante técnicas de minería de datos, particularmente a través de algoritmos de clasificación, las principales características que influyen en los distintos resultados del test de Apgar a los cinco minutos (Apgar5) para los recién nacidos en el Hospital Justo José de Urquiza, de la provincia de Entre Ríos, Argentina.

2. Descripción del Problema

El Descubrimiento de conocimiento en bases de datos, KDD (del inglés Knowledge Discovery in Databases), se ha definido como el proceso no trivial de identificación de patrones válidos, nuevos, potencialmente útiles y comprensibles, a partir del análisis de los datos [4]. El KDD consta de una serie iterativa de etapas o fases, estas son: preparación de los datos que incluye la selección, limpieza, transformación y proyección de los mismos; la etapa de minería de datos; y finalmente la evaluación, y validación del

conocimiento extraído. Para llevar a cabo las tareas de preparación de los datos y aplicación de los algoritmos de minería se requieren conocimientos de estadística, informática, computación, entre otros; pero en la fase de consolidación de ese conocimiento extraído es muy importante la participación de expertos del dominio del problema [5].

La fase de Minería de Datos es la más característica del proceso de KDD, y su objetivo es producir nuevo conocimiento mediante de la construcción de un modelo basado en los datos recopilados [6]. Las técnicas empleadas para la construcción del modelo dependerán del objetivo de la tarea de minería y la información obtenida en etapas previas. Las principales tareas se pueden clasificar como *predictivas*, que resuelven problemas regresión y clasificación mediante técnicas como árboles de decisión, redes bayesianas, redes neuronales, SVM; y *descriptivas* que abarcan problemas de reglas de asociación, correlaciones, clustering, detección de anomalías [7].

El problema de clasificación, puede enunciarse como: dado un conjunto de instancias de entrenamiento y sus etiquetas asociadas, determinar la clase de etiqueta para una instancia de prueba sin etiquetar [8]. Así, el problema de clasificación segmenta en tantos grupos la instancia de prueba como etiquetas de clases definidas haya. Los algoritmos de clasificación típicamente contienen dos fases: una de entrenamiento o aprendizaje, donde se construye el modelo en base a los datos etiquetados, y una etapa de prueba donde se asignan clases o etiquetas a nuevos datos para validar el modelo. Dentro de la tarea de clasificación los Árboles de Decisión son métodos muy utilizados por los algoritmos ya que los modelos resultantes son fáciles de interpretar, su implementación es simple y generalmente son tolerantes al ruido presente en los datos. Distintas técnicas de minería de datos se han aplicado ampliamente en el campo de la salud para obtener modelos que permitan estimar con un cierto grado de confianza el valor de un atributo de clase que resulte útil conocer por su dificultad de obtención o por representar riesgos para la vida humana. A través de los atributos registrados en las historias clínicas de pacientes y sus correspondientes diagnósticos, es posible realizar predicción de enfermedades en base a un conjunto de características presentes en los pacientes. En [9] se analizan los datos de los expedientes clínicos de pacientes prediabéticos con el objetivo de obtener patrones de comportamiento; en [10] se han utilizado técnicas de minería de datos para la vigilancia epidemiológica de la AH1N1, a través de la detección de comportamientos epidemiológicos anormales, con base en las características sociodemográficas y sin-

tomatológicas de los usuarios del servicio de salud y en [11] para el estudio de la mortalidad en México, con el fin de obtener reglas o patrones sobre las características de la defunción y detectar así grupos vulnerables.

2.1. Caso de estudio

En el año 1952 Virginia Apgar [12] propuso un método para evaluar el estado físico del recién nacido inmediatamente después del parto. Esta prueba, denominada test de Apgar, incluye la valoración de cinco características: frecuencia cardíaca, esfuerzo respiratorio, tono muscular, irritabilidad refleja y color. El profesional encargado de la atención del neonato es quien evalúa estas características y asigna un valor de 0 a 2 a cada una de ellas. La puntuación total del test resulta de la suma de los valores dados a los cinco componentes. Esta evaluación se realiza al primer minuto luego del nacimiento y se repite en el quinto minuto de vida del bebé.

Cuando el test de Apgar arroja un resultado de 7, 8, 9 o 10 es normal y es una señal de que el recién nacido presenta un buen estado de salud. Cuando el puntaje es inferior a 7 es una señal de que el bebé necesita atención médica; cuanto más bajo es el puntaje, mayor ayuda necesita el bebé para adaptarse fuera del vientre materno. Un puntaje bajo de Apgar, especialmente a los cinco minutos de vida, es indicativo del compromiso del recién nacido [2], y a menudo se asocia con un mayor riesgo de muerte neonatal [13]. También hay evidencia de que el puntaje de Apgar al quinto minuto es capaz de predecir los resultados adversos del neurodesarrollo a largo plazo, incluidas las puntuaciones reducidas de cociente intelectual y el deterioro motriz [14]. El puntaje de Apgar ha sido desde entonces un método universalmente aceptado para informar la condición del recién nacido y su respuesta a cualquier medida de resucitación.

El hospital Justo José de Urquiza es el único hospital público de la ciudad de Concepción del Uruguay, provincia de Entre Ríos, y cuenta con un promedio de nacimientos entre 1000 y 1200 niños por año [15]. El hospital administra la información de las embarazadas atendidas por medio del Sistema Informático Perinatal (SIP) provisto por el Centro Latinoamericano de Perinatología y Desarrollo Humano de la Organización Panamericana de la Salud.

El SIP posee la funcionalidad de la Historia Clínica Perinatal que, mediante más de 330 variables, permite almacenar la información relacionada con la madre, como sus datos censales, estado de salud, antecedentes familiares, análisis practicados durante el embarazo, información del momento del parto y algunos datos

del recién nacido hasta el momento del alta médica, incluyendo el resultado del test de Apgar realizado a los cinco minutos posteriores al nacimiento.

Analizar los factores que pueden influir en el resultado de este test resulta un problema complejo ya que los datos suelen presentar una alta dimensionalidad, por la cantidad de variables registradas que podrían presentar alguna relación, y generalmente suelen estar muy desbalanceados, dado que la mayoría de los recién nacidos obtienen valores del test Apgar5 mayores a 6.

El objetivo de este estudio es, por medio de algoritmos de clasificación, identificar los factores que más influyen en los resultados del test de Apgar5 realizado a los recién nacidos entre los años 2009 y 2017 en el Hospital Justo José de Urquiza.

3. Materiales y Métodos

Para llevar a cabo el proceso de Minería de Datos se implementó la metodología CRISP-DM ya que proporciona un enfoque sistemático y detallado sobre las tareas y actividades a ejecutar en cada etapa. Para el análisis de la información y la construcción del modelo de clasificación de los factores de riesgos se utilizó la herramienta de software libre WEKA, desarrollado por la Universidad de Waikato para el aprendizaje automático y la minería de datos.

3.1. Recopilación y selección

Como se menciona en la sección anterior, la HCP es un instrumento diseñado para registrar los datos clínicos individuales de la mujer embarazada durante el control prenatal, el parto y el puerperio, y del neonato. El contenido de la HCP se organiza en secciones, que fueron analizadas para determinar la influencia que podrían tener cada una de las variables allí contenidas en el resultado del test de Apgar5. Los criterios de selección empleados se basaron en: el momento en que acontece el hecho registrado, el grado de completitud de la variable y la importancia del nivel de detalle.

La información se encuentra almacenada en una base de datos Microsoft Access y estructurada en cinco tablas diferentes. Mediante una consulta SQL se seleccionaron las variables elegidas, renombrándolas y recodificando sus valores, ya que se almacenan con códigos propios, por ejemplo cuando una madre presenta antecedentes de Hipertensión Arterial se registra en un atributo denominado VAR_0025 el cual fue renombrado a HTA_Pers, y como los valores posibles de dicho atributo se almacena con 'A' y 'B', fueron convertidos en 'S' y 'N' para facilitar luego su interpretación.

3.2. Preprocesamiento

Antes de aplicar una técnica de minería de datos, se deben realizar tareas de preprocesamiento que permitan transformar los datos originales a una forma más adecuada para ser usados por el algoritmo.

La población inicial constaba de 5473 casos de embarazos registrados en el período comprendido entre enero de 2009 y noviembre de 2017. Se excluyeron los registros que no poseían más que el identificador del caso y aquellos que no poseían el valor de Apgar5, resultando un total de 5374 registros (98% de la base original). Se analizó el grado de completitud de las variables extraídas, y en base a eso se descartaron aquellas dimensiones que no superaban el 50% de valores completos. Entre ellas se descartaron algunas que hubieran sido de interés como: peso y talla de la madre, infección puerperal, inmunización, entre otras.

Se unificaron variables por medio de la agregación ya que indicaban un mismo fenómeno medido en distintas etapas del embarazo obteniendo de este modo una sola variable que generalice el evento. Por ejemplo, las pruebas de Sífilis se registran en variables separadas, perteneciendo una de ellas al resultado del análisis realizado durante las primeras 20 semanas de gestación y la otra variable al realizado durante las semanas restantes, ambas categóricas con valores 'S' o 'N'. Para este tipo de casos se creó una variable adicional que se estableció como 'S' si en alguno de los dos análisis la prueba dió positiva y como 'N' si en ambos resultó negativa. El mismo tipo de transformación se realizó sobre dimensiones como: pruebas de hemoglobina inferiores a 11,0 g/dl, toxoplasmosis, Glucemia basal igual o mayor a 105 mg/dL, Bacteriuria, entre otras.

Se discretizaron variables numéricas en intervalos nominales, entre las más importantes se destacan la edad, transformada en intervalos de 5 años; y la variable que representa el valor del test de Apgar5 quedando representada en los siguientes intervalos: 0-3, 4-6, 7-10, considerando el hecho de las diferencias existentes en ciertos rangos de valores, es decir que la medida del riesgo no es lineal.

La vista minable se conformó de 5374 registros de embarazadas y las dimensiones de la misma se describen en la Figura 1 junto al porcentaje de completitud de las mismas.

3.3. Exploración y selección de los datos

Dada la alta dimensionalidad de la vista minable se realizó una selección de atributos basada en la capacidad predictiva de cada una de las dimensiones y

Nombre de la variable	% de Completitud	Nombre de la variable	% de Completitud
Abortos previos	92,0%	Fumadora pasiva 1er.	94,2%
Alcohol 1er.	97,4%	Fumadora pasiva 2do.	92,6%
Alcohol 2do.	95,5%	Gestas previas	100,0%
Alcohol 3er.	94,5%	Glucosa > 1.05	98,0%
Alfabeta	97,3%	Hemorragia 1er. trim.	68,7%
Amenaza parto pretérmino	68,8%	Hemorragia 2do. trim.	68,7%
Analgésia	98,0%	Hemorragia 3er. trim.	68,6%
Anemia	91,8%	Hipertensión (familiares)	97,8%
Anestesia general	98,0%	Hipertensión (personales)	97,8%
Anestesia local	97,0%	Hospitalización	96,6%
Anestesia regional	98,0%	HTA inducida	68,9%
Antecedentes gemelares	92,5%	HTA previa	68,9%
Antibióticos	98,1%	Infección ovular	68,8%
Antitubérola	94,1%	Infección urinaria	69,0%
Antitétánica vigente	94,9%	Infertilidad (personales)	97,2%
Bacteriuria	90,5%	Inicio parto	98,2%
Boca arriba al nacer	91,7%	Ligadura cordón precoz	98,1%
Cardiopatía	68,7%	Nefropatía	68,7%
Cardiopatía antecedentes	97,1%	Ocitócicos en TDP	98,0%
Cirugía (personales)	97,5%	Ocitócicos prealumbramiento	97,2%
Consultas prenatales	95,6%	Otra condición grave	68,7%
Corticoides	97,5%	Otro Antec. familiar	97,3%
Desgarros	98,2%	Otro Antec. Personal	97,0%
Diabetes	68,8%	Otros durante el Parto	93,2%
Diabetes (familiares)	98,0%	Placenta completa	98,8%
Diabetes (personales)	97,8%	Placenta retenida	97,1%
Drogas 1er.	97,3%	Posición parto	98,8%
Drogas 2do.	95,5%	Preeclampsia	68,8%
Drogas 3er.	94,5%	Presentación situación	98,2%
Eclampsia	68,7%	RCIU	68,8%
Eclampsia (familiares)	97,3%	RN Peso para Edad Gestacional	97,6%
Eclampsia (personales)	97,4%	Ruptura de membranas	97,6%
Edad gestacional al parto	97,3%	Prueba sífilis Treponémica	98,8%
Edad gestacional RN	98,5%	Tamaño fetal acorde	97,8%
Edad materna	99,1%	Tamiz. Anten. Chagas	96,5%
Embarazo anterior < 1	100,0%	Tamiz. Anten. Estreptococo B	92,8%
Embarazo múltiple	99,4%	Tamizaje Antenatal - Malaria	92,7%
Embarazo planeado	93,4%	Tamiz. Anten. Toxoplasmosis	94,3%
Enfermedades	95,7%	TBC (familiares)	98,0%
Episiotomía	98,7%	TBC (personales)	97,9%
Estado civil	96,2%	Terminación	99,3%
Estudios	96,6%	Trabajo de parto, detalles	87,9%
Etnia	98,0%	Transfusión	98,0%
Fumadora activa 1er.	97,9%	Tratamiento Sífilis	92,2%
Fumadora activa 2do.	95,9%	Tres abortos espont. Consec.	100,0%
Fumadora activa 3er.	94,9%	Vive sola	93,5%

Figura 1. Variables de la vista minable y completitud del dato

su grado de redundancia común. El objetivo de aplicar esta clase de técnicas es obtener un subconjunto de los datos originales que poseen un nivel expresivo similar al original [16]. Las técnicas empleadas para la selección de atributos se pueden clasificar en base a la estrategia que utilizan para evaluar los atributos, siendo denominadas algunas de tipo *de filtros*, en la que los atributos son evaluados independientemente del algoritmo de aprendizaje y otras de tipo *wrappers* (envoltorios) que evalúan los atributos mediante el uso de estimaciones de la precisión que aporta el subconjunto de datos a un algoritmo de aprendizaje [17]. Para llevar a cabo la reducción de la dimensionalidad se ejecutaron los siguientes algoritmos de selección de atributos disponibles en el software Weka:

- CfsSubsetEval (CFS): este evaluador maximiza la correlación con la clase y minimiza la intercor-

relación del subconjunto.

- ClassifierAttributeEval: evalúa el valor de un subconjunto de atributos considerando la capacidad predictiva individual de cada atributo junto con el grado de redundancia entre ellos [18].
- ConsistencySubsetEval: elige el subconjunto de atributos por el nivel de consistencia respecto a la clase cuando las instancias de entrenamiento se proyectan en el subconjunto de atributos [19].
- CorrelationAttributeEval: evalúa el valor de un atributo midiendo la correlación de Pearson entre él y la clase.
- GainRatioAttributeEval: evalúa cada atributo midiendo su razón de beneficio con respecto a la clase.
- InfoGainAttributeEval: evalúa los atributos midiendo la ganancia de información de cada uno con respecto a la clase. Anteriormente discretiza los atributos numéricos.
- OneRAttributeEval : evalúa la calidad de cada atributo utilizando el clasificador OneR, que emplea el atributo de mínimo error para predecir, discretizando los atributos numéricos.
- ReliefFAttributeEval: selecciona en el conjunto de datos k vecinos, los mas cercanos de la misma clase y de las clases diferentes respectivamente. Calcula la media de la distancia euclídea entre los atributos de estas instancias y asigna un peso a los atributos que luego servirá para determinar su relevancia [20].

En la Tabla 1 se muestran los diez primeros atributos resultantes de aplicar cada uno de los algoritmos de selección de a la vista minable completa. Se puede observar que existen atributos que han sido seleccionados por más de dos técnicas como por ejemplo la edad gestacional del recién nacido, si el bebe nace boca arriba y si la madre fue víctima de violencia durante el 3er trimestre. Sin embargo estas técnicas han descartado atributos que hubieran sido de gran interés clínico tales como: si existieron restricciones de crecimiento intrauterino (RCIU), si el peso del niño es adecuado para su edad gestacional, hipertensión, glucemia, entre otras.

Para unificar los resultados de la ejecución de estas técnicas de selección de atributos se elaboró una tabla en base al resultado arrojado por los mismos. Se calculó la frecuencia de aparición de cada atributo en los resultados y se seleccionaron las 28 dimensiones mas frecuentes de la vista minable que se presentan en la Tabla 2, sin considerar el atributo de clase.

En la vista reducida se analizó la distribución de las clases y, como se mencionó anteriormente, el conjunto de datos es altamente desbalanceado respecto a las

Tabla 1. Primeros 10 atributos escogidos por cada algoritmo de selección

Algoritmos	Atributos
Cfs Subset	eclamp_pers, violencia_3er, etnia, emb_planeado, malaria_antenatal, emb_multiple, episotomia, transfusion, boca_arriba_rn, peso_al_nacer
Classifier Attribute	peso_segun_edad_gest, violencia_1er, violencia_3er, antirubeola, antitetanica_vigente, violencia_2do, alcohol_3er, edad_gest_rn_semanas, alcohol_2do, drogas_2do
Consistency Subset	terminacion_parto, hta_flia, eclamp_flia, fuma_activa_3er, antirubeola, antitet_vigente, rango_edad, estudios, toxoplasmosis, anemia
Correlation Attribute	boca_arriba_rn, edad_gest_rn, semanas, peso_al_nacer, eclamp_pers, emb_multiple, ocitocicos_tdp, ligadura_cordon, amenaza_parto_preter, hospitalizacion, anestesia_regional
Gain Ratio Feature	boca_arriba_rn, peso_al_nacer, edad_gest_rn_semanas, eclamp_pers, malaria_antenatal, violencia_1er, violencia_3er, violencia_2do, transfusion, alcohol_2do
InfoGain Attribute	edad_gest_rn_semanas, peso_al_nacer, boca_arriba_rn, rango_edad, ocitocicos_tdp, episotomia, hospitalizacion, ocitocicos_prealum, ligadura_cordon, corticoides
OneR Attribute	boca_arriba_rn, peso_segun_edad_gest, violencia_3er, antirubeola, violencia_1er, antitet_vigente, violencia_2do, alcohol_3er, abortos_prev, drogas_3er
ReliefF Attribute	anestesia_regional, inicio_parto, terminacion_parto, rango_edad, estudios, boca_arriba_rn, fuma_pasiva_3er, fuma_pasiva_2do, toxoplasmosis, ocitocicos_tdp

Tabla 2. Atributos escogidos en base a su frecuencia de aparición

Atributos	Atributos
alcohol_2do	estudios
alcohol_3er	hospitalizacion
amenaza_parto_preter	ligadura_cordon
anestesia_regional	malaria_antenatal
antirubeola	ocitocicos_tdp
antit_vigente	peso_segun_edad_gest
boca_arriba_rn	rango_edad
drogas_2do	terminacion_parto
drogas_3er	toxoplasmosis
eclamp_flia	transfusion
eclamp_pers	violencia_1er
edad_gest_rn	violencia_2do
emb_multiple	violencia_3er
episotomia	

clases a predecir, lo que se observa en la Figura 2. La clase que representa los valores mas favorables (7-10) de Apgar5 se encuentra presente en el 99,2% de las instancias, mientras que del resto de las clases representan sólo un 0,8% de las instancias.

Las principales técnicas [21] para disminuir el des-

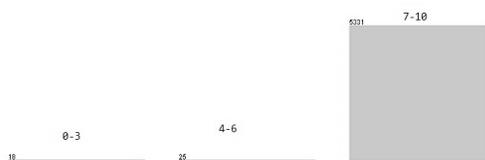


Figura 2. Distribución de las Clases de Apgar5

balanceo se basan principalmente en dos estrategias: seleccionar una mayor cantidad de instancias de datos pertenecientes a las clases minoritarias o descartar instancias pertenecientes a la clase predominante. Para mitigar los efectos adversos de las clases desbalanceadas en la construcción del modelo de clasificación se aplicaron sobre la vista reducida métodos previos de balanceo de clases que permitan obtener mejores resultados en la construcción del modelo. El problema de desbalanceo de los datos tiene una considerable influencia en el desempeño de los algoritmos de clasificación, por lo que resulta conveniente aplicar métodos que resuelvan este problema.

3.4. Entrenamiento y Prueba o Construcción del Modelo

Se experimentó con la vista minable reducida y tres algoritmos de clasificación. Previo a la etapa de entrenamiento se aplicaron a la muestra sin balancear tres técnicas de balanceo diferentes disponibles en Weka:

- **SpreadSubSample**: realiza un re muestreo permitiendo cambiar las proporciones de las distintas clases del conjunto de datos original
- **Class Balancer**: aplica pesos distintos a las clases, haciendo que el algoritmo penalice más los errores de la clase con menor representación
- **SMOTE**: genera nuevas instancias de la clase minoritaria interpolando los valores de las instancias minoritarias más cercanas a una dada.

Se experimentó con los siguientes algoritmos de clasificación del tipo Árbol de Decisión:

J48: Weka implementa el algoritmo C4.5 que es una versión mejorada del J48. Consiste en la construcción del árbol de clasificación a partir del conjunto de entrenamiento y luego realiza una poda (*prunning*) basada en el test de hipótesis para decidir si expande cada rama. En cada nodo selecciona el atributo con mayor *ratio* de ganancia de información, evitando así favorecer la elección de variables con un mayor número de valores.

Random Tree: este algoritmo construye árboles al azar a partir de una combinación de árboles posibles. En este contexto significa que cada árbol tiene la misma posibilidad de ser probado.

REP Tree: construye un árbol de decisión usando la ganancia de información y realiza una poda de error reducido. Solamente ordena una vez los valores de los atributos numéricos. Los valores ausentes se manejan dividiendo las instancias correspondientes en segmentos.

En la Figura 3 se puede observar la interfaz weka con el experimento completo. En la esquina superior derecha se señalan las etapas diseñadas para cada combinación de algoritmo de clasificación y tipo de balanceo de clase.

4. Resultados

Para medir los resultados del aprendizaje generalmente se utiliza la métrica de porcentaje de aciertos, TP Rate, pero dada la asimetría en los datos en estudio se consideró una métrica adicional que combina la precisión y la sensibilidad, F-Measure.

La medida TP Rate está definida por el cociente entre el número de ejemplos que clasifican correctamente para una clase y el número total de ejemplos para la clase estudiada, es decir, la proporción de elementos que están bien clasificados dentro de la clase respecto a todos los elementos que realmente son de la clase.

En el caso en estudio es importante que la predicción de los nacimientos con nivel bajo de Apgar5 sean lo mas certeras posibles, es decir, alta precisión y a su vez que el número de estas predicciones sea razonable, que lo indica la sensibilidad. Por esto se decidió emplear la métrica F-Score que combina la precisión y la sensibilidad para una clase. La F-Score o F-Measure, se considera como una media armónica y puede calcularse con la fórmula:

$$F - Score = \frac{2 * Precision * Sensibilidad}{Precision + Sensibilidad} \quad (1)$$

La precisión representa el porcentaje de casos positivos predichos correctamente mientras que la sensibilidad corresponde a la capacidad para detectar una condición correctamente.

El rango de valores de la métrica de clasificación F-Score oscila entre 0, indicando que el clasificador no es bueno y 1 en el mejor de los casos.

En la Tabla 3 se presentan los resultados de la ejecución de los tres algoritmos de clasificación combinados con cada uno de los tres métodos de balanceo y con la muestra sin balancear. Las métricas establecidas se estudiaron por clases, dada la importancia de que el clasificador construido sea capaz de identificar las clases minoritarias y no solo que obtenga un buen resultado en el promedio general. Como se puede observar, tanto en los valores de TP Rate como de F-Score, la combinación que arrojó mejores resultados es el algoritmo de clasificación Random Tree con un balanceo de clases mediante el algoritmo Class Balancer.

Tabla 3. Resultados de la ejecución por algoritmo y tipo de balanceo

Balanceo de Clases	Clase	TP Rate			F-Score		
		J48	Tree	REP Tree	J48	Tree	REP Tree
ClassBalancer	0-3	100%	100%	78%	0,60	1,00	0,06
	4-6	100%	100%	56%	0,37	1,00	0,25
	7-10	98%	100%	91%	0,99	1,00	0,95
SMOTE	0-3	11%	100%	11%	0,19	1,00	0,20
	4-6	0%	92%	4%	-	0,96	0,08
	7-10	100%	100%	100%	1,00	1,00	1,00
Spread Subsample	0-3	22%	100%	0%	0,35	0,97	-
	4-6	4%	96%	0%	0,07	0,98	-
	7-10	100%	100%	100%	1,00	1,00	1,00
Sin Balanceo	0-3	22%	89%	11%	0,35	0,94	0,19
	4-6	4%	92%	0%	0,07	0,92	-
	7-10	100%	100%	100%	1,00	1,00	1,00

Respecto al algoritmo REP Tree no presenta resultados prometedores con ninguno de los métodos de balanceo, priorizando siempre a la clase mayoritaria.

En base al desempeño de los algoritmos se decidió aplicar Random Tree con el método de balanceo Class-Balancer, ya que esta combinación reportó mayores valores tanto de TP Rate como de F-Score. De la aplicación de estas técnicas, con los parámetros por defecto, resultó la matriz de confusión de la Tabla 4, donde se puede observar para cada clase la cantidad de instancias clasificadas correctamente como tales. Si se observa la diagonal de la matriz se puede ver cómo la totalidad de las instancias fueron clasificadas correctamente. Esto puede deberse a un sobre entrenamiento del modelo.

Tabla 4. Matriz de confusión resultante al aplicar Random-Tree con ClassBalancer

a	b	c	
18	0	0	a= 0-3
0	25	0	b = 4-6
0	0	5331	c = 7-10

Para reducir el árbol de decisión resultante se limitó la profundidad a cuatro niveles, de esta manera se perdió eficacia en la clasificación de la clase mayoritaria, como se observa en la Tabla 5 pero se obtuvo un árbol simple de leer, Figura 4.

Del análisis del árbol de decisión se pueden inferir reglas que conducen a un probable valor de Apgar del recién nacido, por ejemplo:

- Para las madres a las que durante el parto se le suministraron oxicóticos, que han estado hospitalizadas durante el embarazo y el bebé nació con bajo peso en una edad gestacional mayor a las 34 semanas, se

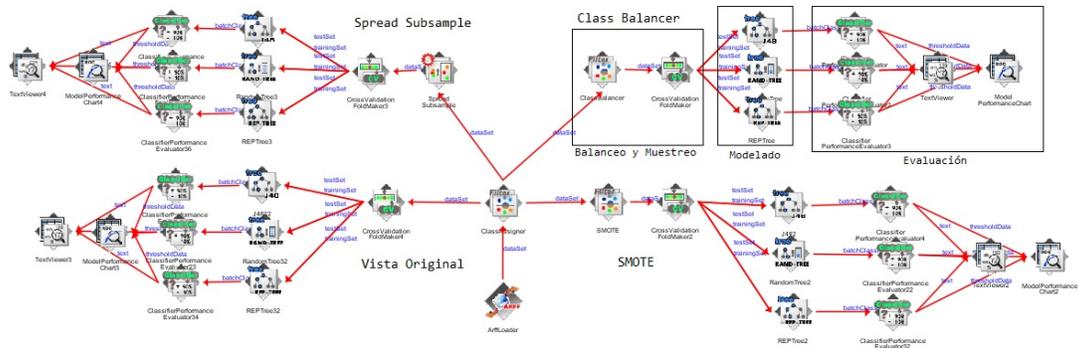


Figura 3. Experimento diseñado para evaluar los algoritmos de clasificación y los métodos de balanceo de clases

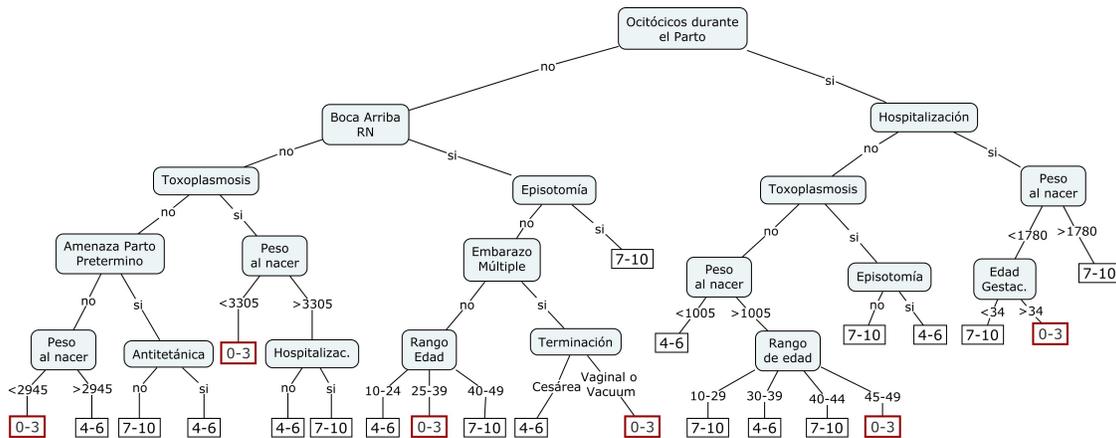


Figura 4. Árbol de decisión resultante de aplicar el algoritmo RandomTree con balanceo ClassBalancer y profundidad de cuatro niveles

Tabla 5. Matriz de confusión resultante limitando la profundidad del árbol

a	b	c	
18	0	0	a= 0-3
0	24	1	b = 4-6
604	1066	3661	c = 7-10

presentan más casos de los valores mas bajos (0-3) de Apgar5.

- Se observan también casos con puntaje bajo de Apgar5 (0-3) en madres a las que no se les suministró oxitócicos durante el trabajo de parto, pero presentaban un embarazo múltiple y el/los niño/s presentaban posición en decúbito dorsal al nacer y el alumbramiento tuvo una terminación Vaginal o por empleo de Vacuum.

- Se observa que en la mayoría de las ramas del árbol el bajo peso del recién nacido es un determinante de los valores más bajos de Apgar.

El árbol construido representa los principales factores que influyen en el valor de Apgar a lo cinco minutos, siendo algunos relativos a las características de la madre y el embarazo (hospitalización, amenaza de parto pretérmino, toxoplasmosis, embarazo múltiple, rango de edad de la madre, antitetánica vigente), otros referidos al momento del parto (oxitócicos durante el parto, episotomía, terminación) y otros propios al recién nacido (peso al nacer, edad gestacional y boca arriba el recién nacido).

Las dimensiones resultantes en este estudio tienen su concordancia en la literatura médica [22]–[24] donde se mencionan como factores influyentes en el Apgar la aplicación de oxitócicos en cualquier momento del trabajo de parto, el bajo peso y la edad gestacional al nacer. Sin embargo, al aplicar técnicas de reducción de dimensionalidad y el algoritmo de clasificación, se descartaron variables de comprobada influencia médica en el estado de salud del recién nacido como

son el peso del niño respecto a su edad gestacional, hipertensión arterial de la madre, restricciones en el crecimiento intrauterino, entre otras.

5. Conclusiones

En este trabajo se realizaron diversos experimentos con el objetivo de clasificar los nacimientos del Hospital Justo José de Urquiza según el valor de Apgar evaluado a los 5 minutos posteriores al alumbramiento. Se aplicaron algoritmos de clasificación, no con el objetivo de predecir valores en futuros nacimientos, sino para estratificar los casos e identificar posibles factores determinantes no estudiados en la actualidad. La técnica de clasificación empleada fue la de Árboles de Decisión debido a que los modelos generados resultan comprensibles para los usuarios finales sin necesidad de poseer conocimiento en técnicas de minería de datos.

En el proceso de KDD una de las tareas que demandó mayor cantidad de tiempo fue la etapa de preprocesamiento de la información, no solo por las transformaciones necesarias en los datos, sino por la dificultad de trabajar con un conjunto altamente desbalanceado y de gran dimensionalidad.

Una de las principales ventajas de utilizar minería de datos respecto al análisis individual de las variables radica en la posibilidad de estudiar en forma conjunta la diversidad de factores que pueden conducir al valor resultante de una clase. En este estudio se han analizado tanto eventos relacionados con los antecedentes clínicos de la madre, hábitos durante el embarazo e información obstétrica del desarrollo del mismo como datos propios del momento del nacimiento, todas en simultáneo. Analizar estas co-ocurrencias podría ser el disparador de nuevos estudios médicos sobre factores que aún no han sido tan explorados.

El paso siguiente a este estudio será la incorporación de la dimensión geográfica de los embarazos a nivel censal con el objetivo de analizar variables propias del ambiente donde vive la madre y realizar minería de datos espacial sobre estos registros.

Referencias

- [1] O. P. d. I. S. OPS, "Información y análisis de salud situación de salud en las américas: Indicadores básicos," *Washington, DC*, vol. 2017, 2017.
- [2] H. A. Thorngern-Jerneck K, "Low 5-minute apgar score: a populationbased register study of 1 million term births," in *Am College Obstetr Gynecol*, 2001, p. 65–70.
- [3] S. Lai and S. Kumar, "Perinatal risk factors for low and moderate five-minute apgar scores at term," *European Journal of Obstetrics and Gynecology and Reproductive Biology*, vol. 2017, 2017.
- [4] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI magazine*, vol. 17, no. 3, p. 37, 1996.
- [5] S. Monserrat and O. Chiotti, "Minería de datos en base de datos de servicios de salud," *Congreso Nacional de Ingeniería Informática Sistemas de Información*, 2013.
- [6] J. Hernández Orallo, C. Ferri Ramirez, and M. J. Ramirez Quintana, *Introducción a la Minería de Datos*. Pearson Prentice Hall, 2004.
- [7] A. De Battista, P. Cristaldo, L. Ramos, J. P. Nunez, S. Retamar, D. Bouzenard, and N. E. Herrera, "Minería de datos aplicada a datos masivos," in *XVIII Workshop de Investigadores en Ciencias de la Computación (WICC 2016, Entre Ríos, Argentina)*, 2016.
- [8] C. C. Aggarwal, *Data Classification: Algorithms and Applications*, 1st ed. Chapman & Hall/CRC, 2014.
- [9] H. J. H. Gómez, "Aplicación de minería de datos a información de pacientes prediabéticos," *Revista Iberoamericana de Producción Académica y Gestión Educativa*, vol. 1, no. 1, 2014.
- [10] R. R. López and C. S. D. W. Clegg, "El cambio climático y sus consecuencias en el desarrollo socio-económico de la ciudad de México," 2010.
- [11] P. G. C. y A. Meza Mendoza, "Estudio de minería de datos para la información de mortalidad en México," *Universidad Nacional Autónoma de México*, 2012.
- [12] V. Apgar, "A proposal for a new method of evaluation of the newborn," *Classic Papers in Critical Care*, vol. 32, no. 449, p. 97, 1952.
- [13] B. M. Casey, D. D. McIntire, and K. J. Leveno, "The continuing value of the apgar score for the assessment of newborn infants," *New England Journal of Medicine*, vol. 344, no. 7, pp. 467–471, 2001.
- [14] E. A. Moore, F. Harris, K. R. Laurens, M. J. Green, S. Brinkman, R. K. Lenroot, and V. J. Carr, "Birth outcomes and academic achievement in childhood: A population record linkage study," *Journal of Early Childhood Research*, vol. 12, no. 3, pp. 234–250, 2014.
- [15] C. del Uruguay, "Indicadores de natalidad en," 2016.
- [16] M. A. Hall, "Correlation-based feature subset selection for machine learning," *Thesis submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy at the University of Waikato*, 1998.
- [17] M. A. Hall and G. Holmes, "Benchmarking attribute selection techniques for data mining," 2000.

- [18] M. A. Hall, "Correlation-based feature subset selection for machine learning," *Thesis submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy at the University of Waikato*, 1998.
- [19] H. Liu, R. Setiono *et al.*, "A probabilistic approach to feature selection—a filter solution," in *ICML*, vol. 96. Citeseer, 1996, pp. 319–327.
- [20] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Machine Learning Proceedings 1992*. Elsevier, 1992, pp. 249–256.
- [21] A. Orriols and E. Bernadó-Mansilla, "The class imbalance problem in learning classifier systems: a preliminary study," in *Proceedings of the 7th annual workshop on Genetic and evolutionary computation*. ACM, 2005, pp. 74–78.
- [22] A. León Pérez and E. Y. Yglesias, "Factores relacionados con el Apgar bajo al nacer," *Revista Cubana de Obstetricia y Ginecología*, vol. 36, pp. 25 – 35, 03 2010.
- [23] A. Laffita B, "Factores que influyen en el Apgar bajo al nacer, en el hospital América de La Habana, CUBA, 2000," *Revista chilena de obstetricia y ginecología*, vol. 70, pp. 359 – 363, 00 2005.
- [24] J. M. Pérez, B. L. Cobián, and C. M. Silva, "Maternal risk factors and premature birth in a public hospital at west of mexico," *Ginecologia y obstetricia de Mexico*, vol. 72, pp. 142–149, 2004.