

APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS EN ATENCIÓN PRIMARIA EN SALUD (APS) PARA EL ANÁLISIS DE RIESGOS EN MUJERES GESTANTES DE LA POBLACIÓN MANIZALEÑA ATENDIDA POR ASSBASALUD

Julián Andrés Gallego¹
Luisa Fernanda Navarro²
Luis Fernando Castillo³

RESUMEN

Objetivo: Determinar las relaciones entre los factores de riesgo existentes en las madres gestantes, a través de las técnicas de minería de datos. **Materiales y Métodos:** Se recolectó una muestra de 15265 registros de mujeres con edades comprendidas entre los 10 y 48 años, datos que pertenecen al período comprendido entre los años 2010 y 2012, incluyéndolos. Los datos se procesaron mediante Microsoft SQL Server 2008 R2 y Microsoft Excel 2007, teniendo en cuenta el proceso KDD y la metodología CRISP DM para la minería de datos usando los algoritmos de k-Means para Clústers, Reglas de Asociación y Métodos de Correlación. **Resultados:** Mediante los algoritmos de clústers se evidenció que las mujeres con antecedentes de aborto componen el 5% de la población estudiada; las mujeres con partos prematuros tienen mayor presencia en edades anormales, siendo este del 2%, mientras que en edades aptas para la gestación la cifra solo es del 1%; el 1% de las mujeres, sin importar la edad, refiere haber tomado licor durante su embarazo; las mujeres que tuvieron abortos y consumieron alcohol fueron el 20%; las mujeres con partos prematuros con consumo de alcohol fueron el 8%. Las mujeres que aceptaron haber consumido drogas fueron el 2% en edad apta

para la gestación y el 3% en edades no aptas, ubicándolas en el clúster que tiene la máxima probabilidad de aborto.

Palabras clave: mujeres gestantes, parto prematuro, aborto, minería de datos, APS.

APPLICATION OF DATA MINING TECHNIQUES IN PRIMARY HEALTH CARE (PHC) FOR RISK ANALYSIS IN PREGNANT WOMEN IN THE POPULATION OF MANIZALES ATTENDED BY ASSBASALUD

ABSTRACT

Objective: To determine the relationship between existing risk factors in pregnant mothers through data mining techniques. **Materials and Methods:** A sample of 15,265 women aged between 10 and 48 years, data pertaining to the period between 2010 and 2012, was collected. Data was processed using Microsoft SQL Server 2008 R2 and Microsoft Excel 2007, taking into account the KDD process and CRISP-DM methodology for data mining using k-Means algorithms for Clusters,

¹ Ingeniero de Sistemas y Telecomunicaciones Assbasalud E.S.E Oficina de Planeación y Sistemas, Assbasalud E.S.E. Manizales, Colombia. juliang373@gmail.com

² Ingeniera de Sistemas y Telecomunicaciones SES Hospital de Caldas. Grupo de Investigación Ingeniería de Software, Universidad Autónoma de Manizales. Manizales, Colombia. lufenapa@gmail.com

³ Doctor en Informática por la Universidad de Salamanca. Facultad de Ingeniería, Universidad de Caldas, Grupo Investigación GITIR. Manizales, Colombia. luisfercastillo@gmail.com

Association Rules and Correlation Methods. **Results:** Using Clustering algorithms it was evident that women with a history of abortion form 5% of the study population; women with premature births have a greater presence in abnormal ages being 2% of the population, while suitable age for gestation is only 1%; 1% of women, regardless of age, reported having consumed liquor during pregnancy; women who had abortions and consumed alcohol were

20%; women with induced labor showed 8% alcohol consumption. Women who admitted having used drugs were 2% in ages suitable for gestation and 3% in ages non suitable for conception, placing them in the cluster having the highest probability of abortion.

Key words: pregnant women, induced labor, abortion, data mining, PHC.

INTRODUCCIÓN

En el momento en que una mujer queda en embarazo es susceptible, junto a su hijo, a fenómenos adversos conocidos como riesgos reproductivos. Mediante la estrategia de Atención Primaria en Salud (APS) llevada a cabo por Assbasalud E.S.E. se intenta reducir, entre muchas otras cosas, los índices de riesgo reproductivo de la población Manizaleña (1, 2).

La Atención Primaria en Salud como tal fue definida en 1978 en la Declaración de Alma-Ata y, después de casi 30 años de experiencia, se puede concluir que los sistemas de salud que cumplen con los principios de APS han logrado mejores resultados sanitarios y han aumentado la eficiencia en la atención de salud tanto individual como colectiva, facilitando un mejor relacionamiento y participación entre usuarios y prestadores de salud públicos y privados (3).

Asimismo, intenta bajar los índices de anomalías funcionales y/o estructurales del feto debidas a factores presentes antes del nacimiento como: aspectos genéticos, ambientales o sociales (1, 3). Assbasalud E.S.E. se ha esforzado por calcular los riesgos a los que una gestante y su producto se exponen. La frecuencia de malformaciones clínicamente importantes en los recién nacidos está en torno del 3%. Sin embargo, los defectos congénitos en general son más frecuentes, afectando entre el 5% y el 10% del total de los nacimientos (3-5). La mortalidad neonatal y de las madres gestantes en las etapas de gestación, así como en los días posteriores, son algunos de

los grandes problemas a los que se enfrentan las estrategias en salud pública (6). Esto obliga a que las instituciones prestadoras de salud adopten medidas para disminuir estos índices negativos y que afectan desfavorablemente a la sociedad. Uno de los puntos de partida es aprender de las experiencias que se han tenido a lo largo de los años (3). Las cifras de mortalidad en cualquiera de los factores del binomio madre-hijo y los defectos congénitos, cobran importancia en todas las instituciones de salud donde se llevan a cabo atenciones perinatales, pues demuestran la calidad del servicio que le están prestando a la sociedad. Por esta razón, la intención de este trabajo es evidenciar cuáles son las condiciones sociales y clínicas que afectan a las madres gestantes y a sus productos, de forma que la extracción de conocimiento contribuya al mejoramiento de la calidad en la prestación del servicio de salud.

Más que contar variables que puedan ser representativas a la hora de hacer estos conteos, se debe hacer un análisis detallado de cuáles son las variables que no solo representen riesgos, sino además cuáles permiten saber de antemano el patrón de riesgo al que se encuentran expuestas las personas, cuando existe un conjunto de riesgos ya identificados como causantes de los eventos adversos. En otras palabras, se requiere de un análisis multivariado; razón por lo cual se esbozó un plan para que se pudiera extraer conocimiento a partir de los repositorios de datos que posee la entidad (3). Este proceso es conocido como KDD y en esta oportunidad está siendo aplicado a las 4651

debido a que no hay una tabla de referencia para cada código. Además, se aplicaron reglas de exclusión a personas sin dato de nivel de estudio, personas sin dato de sospecha de maltrato y personas sin dato de trastorno mental. No se incluyeron datos de identificación del paciente debido a que se respeta la confidencialidad de la información de Historia clínica. El objetivo de la fase de construcción e integración es construir un repositorio de datos, en este caso mediante una vista de Base de Datos, en donde se estructuran de forma tal que cada registro sea una etapa de gestación diferente. Así, toda la información recolectada se encuentra en vistas diseñadas en el SQL Server 2008 R2, lo cual las define por sí mismas como bodegas de datos. Los datos extraídos y transformados se pueden visualizar a través de las vistas generadas anteriormente, de esta forma quedan disponibles para ser cargados y consumidos a través de Microsoft Excel, ya que en este caso es el software encargado de hacer la interfaz gráfica, no solo de consumo sino también de la generación del modelo de minería de datos y la visualización de los resultados arrojados por el SQL Analysis Server.

La fase de modelado en la minería de datos es donde se seleccionan las técnicas de minería a emplear y se generan patrones y tendencias a partir de los datos y, a continuación, son usados para realizar análisis adicionales o predicciones (10). Este procedimiento se realiza mediante el Cliente de minería de datos para Excel, proporcionando asistentes que facilitan la creación de modelos de minería (11).

Modelo de clústers

La primera técnica seleccionada genera un modelo de clústers. El algoritmo empleado utiliza técnicas iterativas para agrupar los casos de un conjunto de datos en clústers que contienen características similares. Estas agrupaciones son útiles para la exploración

de los datos, la identificación de anomalías en ellos, y la creación de predicciones. Estos modelos identifican relaciones en un conjunto de datos que no podrían derivar lógicamente a través de la observación casual. El algoritmo de clustering entrena el modelo estrictamente con las relaciones que existen en los datos y con los grupos que identifica en el proceso (12).

Modelo de asociación

La segunda técnica seleccionada genera un modelo de asociación que detecta formas de agrupamiento entre elementos que aparecen juntos en un mismo momento: en el tipo más común de modelo de asociación, las variables booleanas, que representan un valor Sí/No o Falta/Existe, se asignan a cada atributo, como un nombre de producto o evento. Para cada conjunto de elementos el algoritmo crea puntuaciones que representan el soporte y la confianza. Estas puntuaciones se pueden usar para clasificar y derivar reglas interesantes de los conjuntos de elementos.

A continuación, el algoritmo cuenta el número de veces que aparece cada conjunto de elementos y calcula la importancia relativa de cada uno de ellos en todas las transacciones.

El algoritmo usa esta información sobre los conjuntos de elementos, para generar reglas que se puedan usar para predecir asociaciones o realizar recomendaciones. A cada recomendación se le asigna una probabilidad en función de la solidez de las asociaciones (13).

Correlación de datos

El Coeficiente ϕ o Coeficiente de correlación de Mathews r_{ϕ} , se emplea cuando se encuentran dos variables dicotómicas, es decir, variables para las cuales solo es posible considerar dos modalidades.

Tabla 1. Análisis de correlación para edades consideradas normales para la gestación.

Variable	ABORTO			Total	Valor
	No	Sí			
Hipertensión	No	2010	120	2130	0.04274103
	Sí	130	2	132	
Riesgo físico	No	1501	82	1583	0.00806097
	Sí	575	34	609	
Riesgo químico	No	1929	107	2036	0.00590101
	Sí	147	9	156	
Riesgo biológico	No	1517	89	1606	0.01846815
	Sí	559	27	586	
Riesgo social	No	406	30	436	0.0353621
	Sí	1670	86	1756	

Área	No	Sí	Total	Valor	
	Anemia	Urbana	1803	98	1901
Rural		273	18	291	0.01561689
Anemia	No	1971	105	2076	0.02319807
	Sí	107	9	118	

En gestantes en edades normales, entendiéndose de 15 a 34 años, al analizar la presencia de abortos (116 casos), y la presencia de mujeres con riesgos físicos (609 casos), encontrándose 34 casos en total con presencia de aborto y riesgo físico al mismo tiempo, muestra una correlación de 0,00806097, lo cual representa que en las gestantes atendidas por Assbasalud E.S.E. se encuentra una muy baja correlación entre la presencia de aborto y la condición de presencia de riesgo físico.

En gestantes en edades normales, entendiéndose de 15 a 34 años, al analizar la presencia de abortos (116 casos), y la presencia de mujeres con riesgos químicos (156 casos), encontrándose 9 casos en total con presencia de aborto y riesgo químico al mismo tiempo, muestra una correlación de 0,00590101, lo cual representa que en las gestantes atendidas por Assbasalud E.S.E. se encuentra una muy baja correlación entre la presencia de aborto y la condición de presencia de riesgo químico.

En gestantes en edades normales, entendiéndose de 15 a 34 años, al analizar la presencia de abortos (116 casos), y la presencia de mujeres con riesgos biológicos (586 casos), se encontraron 27 casos en total con presencia de aborto y 161 con

riesgo biológico al mismo tiempo, mostrando una correlación de 0,01846815, lo cual representa que en las gestantes atendidas por Assbasalud E.S.E. se encuentra una muy baja correlación entre la presencia de aborto y la condición de presencia de riesgo biológico.

En gestantes en edades normales, entendiéndose de 15 a 34 años, al analizar la presencia de abortos (116 casos), y la presencia de mujeres con riesgos sociales (1756 casos), encontrándose 86 casos en total con presencia de aborto y riesgo social al mismo tiempo, muestra una correlación de 0,0353621, lo cual representa que en las gestantes atendidas por Assbasalud E.S.E. se encuentra una muy baja correlación entre la presencia de aborto y la condición de presencia de riesgo social.

En gestantes en edades normales, entendiéndose de 15 a 34 años, al analizar la presencia de abortos (116 casos), y la presencia de mujeres con anemia (114 casos), encontrándose 9 casos en total con presencia de aborto y anemia al mismo tiempo, muestra una correlación de 0,02319807, lo cual representa que en las gestantes atendidas por Assbasalud E.S.E. se encuentra una muy baja correlación entre la presencia de aborto y la condición de presencia de anemia.

Evaluación de los datos

Las mujeres con antecedentes de aborto componen el 5% de la población estudiada.

Estas cifras son similares tanto para mujeres con edades aptas para la gestación, como para mujeres con edades no aptas. Esta cifra es superior al promedio que se encuentra, del 3% tanto para Colombia como para Latinoamérica.

Las mujeres registradas con partos prematuros tienen mayor presencia en edades anormales, siendo este del 2%, mientras que en edades aptas para la gestación la cifra solo es del 1%.

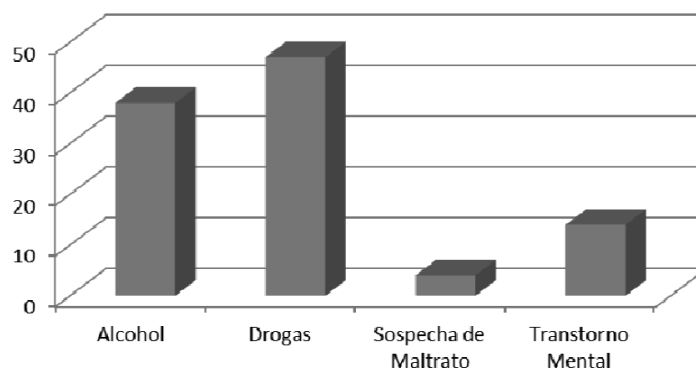


Figura 2. Consumo de drogas y alcohol durante la gestación.

Algunos de los diagnósticos que se encontraron a las mujeres que alguna vez fueron captadas como gestantes por Assbasalud E.S.E., no cuentan con unificación de criterios médicos, debido a la subjetividad que por su naturaleza representa. Por esta razón, se presentan 62 diagnósticos de partos prematuros en la población afectada. Cabe aclarar que, debido al nivel de atención que presta la entidad, toda madre gestante que sea detectada con alguna alteración en su etapa de gestación, es remitida a otra institución que preste servicios con niveles superiores de atención, lo cual hace que este tipo de diagnósticos no sea superior a lo anteriormente descrito.

El 1% de las mujeres, sin importar la edad, refiere haber tomado licor durante su embarazo. El consumo de alcohol durante el embarazo se ha asociado con defectos físicos y psíquicos al nacer la criatura, que reciben el nombre de síndrome alcohólico fetal y se presenta en el 6% de los bebés, con madres que abusaron del alcohol durante la etapa de gestación. Sin embargo, la cifra de las mujeres que tuvieron abortos y consumieron alcohol fue del 20%, y para las mujeres con partos prematuros con consumo de alcohol fue del 8%. Estos hallazgos se presentaron en mujeres comprendidas entre los 15 y 35 años, edades consideradas aptas para la gestación.

CONCLUSIONES

El correcto uso de las herramientas informáticas puede ayudar a identificar o descubrir relaciones con el objetivo de establecer, o proponer estrategias para contrarrestar problemas de salud o prevenir futuras complicaciones, contribuyendo así a la mejora en la calidad de vida de la población.

La modelación de los datos obtenidos en los estudios realizados ha permitido a Assbasalud E.S.E. concentrar los esfuerzos en el 5% de la población estudiada susceptible de presentar abortos.

Según los resultados obtenidos, entre las mujeres encuestadas, quienes aceptaron haber consumido drogas fueron el 2% en edad apta para la gestación y el 3% en edades no aptas, ubicándolas en el clúster que tiene la máxima probabilidad de aborto, así pues, se deben implementar estrategias como los principios de APS que han hecho que la eficiencia en la atención aumente.

Assbasalud E.S.E. se ha esforzado en identificar los riesgos a los que una madre gestante y su hijo se exponen, pues las malformaciones en los recién nacidos rodean el 3% y los defectos congénitos afectan entre el 5% y el 10%, todo

esto debido a factores presentes en el nacimiento como aspectos genéticos, ambientales o sociales.

Toda esta información sirve para hacer un llamado a las instituciones prestadoras de salud para concentrar más esfuerzos en mejorar la atención a las mujeres, de tal forma que se promuevan las consejerías de familia, los programas de vacunación, asistencia odontológica, participación en programas comunitarios educativos y promoción de lactancia materna, con el fin de que la estrategia de Atención Primaria en Salud rinda frutos positivos en el menor tiempo posible.

Trabajos futuros

Esta herramienta debe ser tenida en cuenta por las directivas institucionales, con el fin de brindar apoyo a la falta de métodos modernos de generación de informes, que dependen de grandes volúmenes de datos recolectados a lo largo de los años y su potencial no ha sido explotado al máximo. Estos métodos de análisis

de datos deberán ser implementados en aquellas entidades que han incorporado Atención Primaria en Salud como una de las estrategias de salud pública y que, al igual que Assbasalud E.S.E., prestan el servicio de obstetricia.

Se seguirá ampliando el proceso de análisis de datos en la entidad con el fin de mejorar los procesos a futuro.

AGRADECIMIENTOS

Agradecimientos muy especiales al Ing. Richard Nelson Román Marín, jefe de Planeación y Sistemas de Assbasalud E.S.E. Al Dr. Luis Fernando Castillo, tutor del Trabajo de Grado. Al Dr. Carlos Londoño, Médico Territorial de Salud de Caldas. Al Dr. Mauricio Alba, Director del programa de Maestría en Gestión y Desarrollo de Software de la Universidad Autónoma de Manizales y, en general, a todas las personas que de una u otra forma se vieron involucradas en dicho proceso.

REFERENCIAS

1. Assbasalud E.S.E. Atención primaria en salud en el marco de las redes integradas de servicios de salud; 2012. Disponible en: <http://www.assbasalud.gov.co/>
2. Malagón G. (2011.). Editorial. Atención Primaria en Salud (APS): Una estrategia para mejorar la cobertura y la calidad. Revista Academia Nacional de Medicina 2011; 83. Disponible en: <http://encolombia.com/medicina/revistas-medicinas/academedicina/vola-83/atencion-primaria-en-salud/>
3. Fescina RH, De Mucio B, Díaz JL, Martínez G, Serruya, S. (2010). Guías para el continuo de la atención de la mujer y el recién nacido focalizadas en APS. Montevideo: Publicación Científica CLAP/SMR N° 1573; 2010. Disponible en: https://docs.google.com/file/d/0B3o7GPnDqjGEUWtuV2hYVzdSM2VHWWktRWZidHo2QQ/edit?usp=drive_web&pli=1
4. Cárdenas V. Introducción al Análisis Service de SQL Server 2008 R2 [audio podcast]; 2011. Disponible en: <http://www.youtube.com/watch?v=4-FxpeOFhHU>
5. Organización Mundial de la Salud. Conferencia Internacional sobre Atención Primaria de Salud patrocinada por la OMS y el UNICEF; párrafo 10; 1978. Disponible en: http://www.who.int/topics/primary_health_care/es/
6. Castaño JJ, Esquivel VM, Ocampo P, Páez ML, Rico LM, Santacoloma V, et al. Características de madres gestantes y sus recién nacidos en relación con la edad de las madres en el departamento de Caldas (Colombia), 2003-2008. Archivos de Medicina (Col) 2011; 11(1):23-38. Disponible en: <http://www.redalyc.org/articulo.oa?id=273819434002>

7. Tercero I. Minería de datos (data mining) [blog]; 2010 24 de febrero. Disponible en: http://coevolucion.net/index.php?option=com_content&view=article&id=143:mineria-de-datos-data-mining
8. Sánchez R, Rosete A, Rodríguez A, Brito R. Empleo de minería de datos en la predicción de diabetes. Informática en salud 2009. Disponible en: <http://www.informatica2009.sld.cu/Members/ardguez/empleo-de-mineria-de-datos-en-la-rediccion-de-diabetes-reprocesado-de-datos/>
9. Vallejos SJ. Minería de Datos. Universidad Nacional del Nordeste, Corrientes, Argentina; 2006. Disponible en: http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/Mineria_Datos_Vallejos.pdf
10. Msdn. Algoritmos de minería de datos (Analysis Services: Minería de datos); 2013. Disponible en: <http://msdn.microsoft.com/es-es/library/ms175595.aspx>
11. Msdn. Complementos de minería de datos de Microsoft® SQL Server® 2012 para Microsoft® Office® 2010; 2014. Disponible en: <http://www.microsoft.com/es-es/download/details.aspx?id=29061>
12. Msdn. Microsoft Algoritmo de clústeres; 2014. Disponible en: <http://msdn.microsoft.com/en-us/library/ms174879.aspx>
13. Msdn. Referencia técnica del algoritmo de asociación de Microsoft; 2008. Disponible en: [http://msdn.microsoft.com/es-es/library/cc280428\(v=sql.100\).asp](http://msdn.microsoft.com/es-es/library/cc280428(v=sql.100).asp)