

Propuesta de una Metodología Utilizando Minería de Datos para Estrategias de Medicina Preventiva Más Eficaces

Ezequiel Alonso Sanz-Moreno, Raquel Torres-Peralta, Federico Cirett-Galán, Mario Barceló-Valenzuela.

Universidad de Sonora, Departamento de Ingeniería Industrial, Rosales y Blvd. Luis Encinas S/N
CP. 83000, Hermosillo, Sonora, México.

a211200120@alumnos.uson.mx, rtorres@industrial.uson.mx,
fcirett@industrial.uson.mx, mbarcelo@industrial.uson.mx

Resumen. Se propone una metodología que incremente de manera significativa la generación y difusión de conocimiento y contenido informativo en una Institución de Salud Pública, para el envío de información personalizada y dirigida, vía medios electrónicos, a sus derechohabientes según características demográficas, fisiológicas y criterios especiales, a la vez que mejore la toma de decisiones en Campañas de Medicina Preventiva enfocadas en Obesidad y Diabetes, esto debido a que actualmente no se cuenta con técnicas o métodos para aprovechar el conocimiento implícito en los registros de consultas e historial médico de sus derechohabientes. Mediante el uso de diversas herramientas de Minería de Datos y Estadísticas se espera poder detectar a aquellos con bajo, medio y alto riesgo de padecer alguna de estas enfermedades, observando tendencias y patrones de riesgo que ayuden a alertar oportunamente a los derechohabientes.

Palabras Clave: Minería de Datos, Generación de Conocimiento, Medicina Preventiva, Análisis de Datos.

1 Introducción

En años recientes la gran mayoría de personas reciben servicios de salud de parte de Instituciones de Salud Públicas y Social, en 2012 en Sonora el 69.1% de la población estaba afiliada a una institución social o privadas y el 30.9% restante contaba con acceso al Seguro Popular [1], por ello el rol que desempeñan en la sociedad es importante [2]. La salud pública tiene por objetivo el fomentar, elaborar y aplicar políticas para proteger el bienestar de la población y prevenir, reducir la probabilidad de aparición e impedir o controlar la progresión de enfermedades [3].

La creciente demanda de servicios en las instituciones de salud pública puede ser vista como una epidemia, ya que se está presentando en un mismo grupo de personas en una

Ezequiel Alonso Sanz-Moreno, Raquel Torres-Peralta, Federico Cirett-Galán, Mario Barceló-Valenzuela, *Propuesta de una metodología utilizando minería de datos para estrategias de medicina preventiva más eficaces*, en: Mario Barceló Valenzuela, Alonso Pérez Soltero, Oscar Mario Rodríguez Elías, Guillermo Valencia Palomo, Ramón René Palacio Cinco, René Daniel Fornés Rivera (Eds.), *Avances de Investigación en Ingeniería en el Estado de Sonora*, pp. 211-217, 2017.

determinada región, donde influyen y comparten características como sexo, edad, síntomas, **ubicación**, entre otras [4]. Esta tendencia se está viviendo en México con algunas enfermedades como la obesidad y diabetes, teniendo implicaciones negativas en la salud, la economía, producción, entre otras áreas [5]. Cabe señalar que México es el primer lugar mundial en obesidad infantil, pero esta tendencia pudiera contrarrestarse mediante la prevención [6].

Una de las prácticas de salud pública para prevenir o disminuir el impacto de enfermedades y epidemias, es la planeación y aplicación de Estrategias de Medicina Preventiva, se pueden definir como “La ciencia y el arte de prevenir enfermedades, prolongar la vida y promover la salud, ejercida con el fin de interceptar las enfermedades en cualquier fase de su evolución” [3].

Al recopilar información como los tipos de enfermedades y las ubicaciones donde se presentan [7], las Instituciones de Salud pueden extraer conocimiento mediante el uso de herramientas de minería de datos para conocer las zonas con mayor incidencia y probabilidad de ocurrencia de padecimientos como, en este caso, obesidad y diabetes, inclusive se puede tener información en tiempo real y actuar para prevenirlas.

Grunseit y otros [8] concluyen en su investigación que la utilización de medios masivos es un muy buen primer paso para lograr la prevención de la obesidad, gracias en mayor parte a que se había generado conocimiento y contenido de calidad mediante el análisis de datos para campañas de prevención. En la actualidad, las redes sociales representan una gran ventaja, pues llegan directamente a segmentos de la población definidos por los administradores de contenido, en este caso resultan una herramienta idónea para la difusión de material informativo.

Con la ayuda de herramientas tecnológicas de análisis de datos, las instituciones de salud pueden implementar campañas de medicina preventiva efectivas que alerten y beneficien a la población de manera más efectiva que con los métodos tradicionales. En esta propuesta se intenta cambiar la manera masiva y generalizada en la que se difunde el material de medicina preventiva y apoyar en las estrategias de prevención y ataque a enfermedades crónicas como diabetes y obesidad con un enfoque segmentado por grupos específicos de acuerdo a sus características. Para esto, la colaboración de un equipo multidisciplinario que cubra las áreas de medicina, ciencia de datos, gestión del conocimiento, diseño gráfico y mercadotecnia, es de gran importancia.

2 Marco Teórico

Las instituciones de salud han registrado datos, como el tipo de enfermedad, decesos, tiempo de propagación, ubicación geográfica, entre otros, de brotes de enfermedades que afectan a millones de personas [7], generando inmensos volúmenes de datos en bruto. Arab [9] menciona que existen sistemas geográficos para ubicar y delimitar zonas de interés, estos incluyen representaciones en forma de mapas e imágenes cartográficas,

Propuesta de una metodología utilizando minería de datos para estrategias de medicina preventiva más eficaces

mismos cuyo uso se ha convertido en componente central para tomar decisiones y acciones preventivas contra enfermedades como la diabetes [10].

Para trabajar con grandes volúmenes de datos existen diversos tipos de herramientas, como Minería de Datos, según Mining y otros [11] mediante la aplicación de técnicas avanzadas de análisis para explorar e identificar útiles e inesperados patrones y reglas en los datos, esta provee de conocimiento sumamente relevante para la predicción de futuros escenarios y disminución del impacto de enfermedades.

2.1 Uso de Minería de Datos en la Salud y Medicina Preventiva

Youssefagha y otros [12] realizaron un estudio en escuelas de Pennsylvania, USA, donde se recopilaron datos como género, grado escolar, raza/etnia, peso, altura, entre otros. Se concluyó que con el uso de técnicas de minería de datos se facilitó el análisis de una gran cantidad de registros, lo que arrojó información sobre la relación existente, y no explícita, entre el IMC y el padecer enfermedades como diabetes, asma, alergias, problemas ortopédicos, y enfermedades cardiovasculares en adolescentes.

Respecto a diabetes tipo II, Su y otros [13] desarrollaron un modelo de minería de datos con herramientas como redes neuronales, árboles de decisión, regresión logística, entre otras, para analizar datos de escaneos antropométrico de personas, llegando a la conclusión de que con este método se puede diagnosticar el padecimiento de diabetes de forma más rápida y económica que con los actuales estudios bioquímicos. Dejando en claro el impacto benéfico de la minería de datos en la aplicación de medicina preventiva una vez que se tienen definidas las variables y pre-filtrados los datos.

3 Entorno de Aplicación

El presente proyecto se desarrollará en una Institución de Salud Pública del Estado de Sonora que presta sus servicios a un aproximado de 182,000 derechohabientes. Debido a la creciente demanda de servicios, a la actual reducción en presupuesto y medidas de austeridad, se está apostando a la predicción y prevención en lugar del tratamiento de futuros padecimientos. Para lograrlo esta Institución cuenta con un Área de Medicina Preventiva que se enfoca en enfermedades como obesidad y diabetes.

Pese a que se cuenta con una aplicación celular, uso de mensajes SMS y una tarjeta de citas “inteligente”, las campañas de medicina preventiva carecen del impacto deseado, pues al tratarse del envío masivo de información sin que exista algún tipo de filtro que discrimine entre los distintos tipos de derechohabientes, se difunde material informativo generalizado, que muchas veces no aplica a la condición de quien lo recibe, provocando que a la larga se ignoren los futuros mensajes y se pierda el interés de seguir las indicaciones recibidas por estos medios para el cuidado de su salud.

Actualmente la Institución cuenta con grandes bases de datos con expedientes médicos y ubicaciones de derechohabientes que han sido reunidos durante más de una década, por desgracia, al no explotar estos recursos, la generación de información de calidad y conocimiento para futuras campañas es casi nula, de modo que el material gráfico (infográficos, flyers, imágenes ilustrativas) y contenido informativo (artículos, posts, tweets o mensajes de texto) no se dirige a los distintos segmentos de la población de acuerdo al nivel de riesgo y características demográficas.

3.1 Formulación del Problema

Esta Institución de salud cuenta con programas de medicina preventiva aplicados a la población en general con el fin de prevenir padecimientos y motivando a seguir ciertas acciones en mejora de la salud. Aunque el trabajo se desempeña en tiempo y forma, se considera que el uso de herramientas para el análisis de datos es clave en la generación de conocimiento, de modo que se puedan analizar y determinar criterios para la difusión de materiales y contenido informativo dirigido especialmente a grupos de derechohabientes que compartan características como edad, sexo, ocupación, ubicación geográfica, entre otras, y así reciban información específicamente de sus padecimientos o conozcan los riesgos que corren de acuerdo a su historial médico y cómo prevenirlos.

4 Propuesta de Solución

Basado en lo antes expuesto y tomando en cuenta las carencias y necesidades de la Institución en cuanto a sus campañas de medicina preventivas y el tipo de información que hace llegar a sus derechohabientes para la toma de acciones tempranas en el cuidado de su salud, con especial énfasis en obesidad y diabetes, la propuesta metodológica se centra en el envío y generación de contenido informativo y conocimiento a través de la explotación de los datos disponible en los registros y seguimiento a pacientes de al menos los últimos 5 años. En primera instancia se pretende generar una visión del estado actual de obesidad y diabetes entre la población atendida en la Institución, analizar el comportamiento en los últimos años y su probable tendencia de acuerdo a predicciones o proyecciones basadas en técnicas de análisis, estadística descriptiva y herramientas de minería de datos. Este análisis se realizará por intervalos constantes de tiempo para evaluar cambios y monitorear el comportamiento de variables.

Después del análisis inicial, se separará a los derechohabientes, comenzando con dos grandes grupos, el primero que contenga a toda aquella persona que ya cuente con un diagnóstico de obesidad o diabetes, y el segundo con todas aquellas propensas a padecer alguna de estas enfermedades. Análisis posteriores permitirán la generación de criterios para refinar la granularidad de los grupos de pacientes con diagnósticos y con potencial de desarrollar alguno de los padecimientos. El objetivo es lograr una segmentación adecuada

Propuesta de una metodología utilizando minería de datos para estrategias de medicina preventiva más eficaces

a cada tipo de persona que comparta características demográficas y geográficas. La revisión de estos grupos se hará cada mes, con el fin de detectar nuevos individuos y ver la necesidad de replantear criterios. En este proceso se utilizarán algoritmos de aprendizaje no supervisado para la detección de grupos.

Una vez que exista la segmentación de derechohabientes, se determinará qué herramientas de minería de datos o combinación de estas, serán las que detecten a los derechohabientes con potencial de desarrollar alguna de estas enfermedades. Con esto se generarán reportes periódicos dirigidos al Área de Medicina Preventiva, para lograr una valiosa contribución a la toma de decisiones y la generación de estrategias de prevención.

Por último se realizará una evaluación para su contraste con el análisis inicial, determinando el beneficio obtenido, así como el nivel y relevancia del conocimiento generado y difundido. En la Figura 1 se observa la metodología antes propuesta.

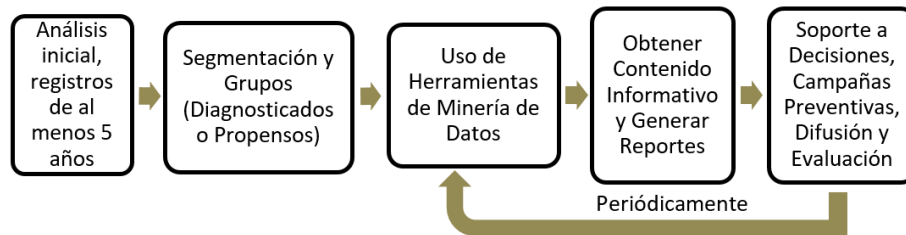


Fig. 16. Propuesta metodológica para generar contenido informativo.

4.1 Reportes para el Área de Medicina Preventiva

La metodología propuesta busca la obtención de información, patrones, tendencias, predicciones, escenarios futuros y conocimiento en general referentes a obesidad y diabetes, esto será presentado mediante una serie de reportes periódicos para el Área de Medicina Preventiva (Figura 2).



Fig. 17. Reportes para el área de Medicina Preventiva de la Institución.

5 Conclusiones

Se espera que el desarrollo de esta metodología incremente significativamente la generación de conocimiento y contenido informativo para la toma de decisiones enfocadas a obesidad y diabetes, también que permita la detección de derechohabientes en riesgo de padecer alguna de estas enfermedades, mediante el análisis de registros de consultas e historial médico, para estructurar estrategias que difundan información personalizada de manera más efectiva y prevenga o reduzca el impacto de dichas enfermedades. Se busca lograr la segmentación de derechohabientes, tomando en cuenta el alto, medio o bajo riesgo de padecer obesidad y diabetes, en caso de ya padecer alguna de estas enfermedades, clasificarlos de acuerdo a su nivel de gravedad, para una mejor toma de decisiones. Además, generar reportes de la situación actual y otros de manera periódica para el Área de Medicina Preventiva con información precisa para su uso en campañas, envío de información vía medios electrónicos e impresos.

Además, como existe una creciente demanda de servicios médicos en la Institución, por enfermedades prevenibles como obesidad o diabetes, se está optando por crear, lanzar y mantener Campañas Preventivas, y una metodología que integre herramientas de Minería de Datos, mismas que han demostrado facilitar y agilizar el procesado, análisis y visualización de registros, permitiendo generar información y conocimiento sólidamente sustentado en el análisis de datos, detección de tendencias, patrones, interrelaciones no evidentes, proyecciones, entre otros, permitirá a la Institución el tomar mejores decisiones en cuanto a la ejecución de dichas campañas y el envío de información para el cuidado y mejora de la salud de manera personalizada y segmentada a sus derechohabientes, según sus características demográficas, región o nivel de riesgo, ayudada mayormente de medios electrónicos.

La integración de un grupo interdisciplinario es necesario para lograr este objetivo.

Referencias

1. Gobierno del Estado de Sonora: Informe a Sonora 2012. Transparencia Sonora. <http://transparencia.esonora.gob.mx/transparencia/Documentos/sonora-saludable.pdf> (2012). Accedido el 14 de Junio de 2017.
2. Olkiewicz, M., Bober, B.: Role of quality in healthcare service provision process. *Prog Heal Sci* 2015. 5, 41–53 (2015)
3. Cascón, M.: Medicina Preventiva y Salud Pública y Comunitaria. *Murcia*. 1–6 (2008)
4. Brémond, P., Roca, Y., Frédérique, S., Walter, A.: Evolution of Dengue Disease and Entomological Monitoring in Santa Cruz. *PLoS One*. 1–21 (2015)
5. Clarke, J.: Preventive medicine: A ready solution for a health care system in crisis. *Popul. Health Manag.* 13, 3–11 (2010)
6. Elder, J.: Mexico and the USA: The world's leaders in the obesity epidemic. *Salud Pública Mex.* 55, (2013)

Propuesta de una metodología utilizando minería de datos para estrategias de medicina preventiva más eficaces

7. Buczak, A., Baugher, B., Babin, S., Ramac-thomas, L., Guven, E., Elbert, Y., Koshute, P., Velasco, J., Roque, V., Tayag, E., Yoon, I., Lewis, S.: Prediction of High Incidence of Dengue in the Philippines. *PLoS Negl. Trop. Dis.* 8, 1–13 (2014)
8. Grunseit, A., O'Hara, B., Chau, J., Briggs, M., Bauman, A.: Getting the message across: Outcomes and risk profiles by awareness levels of the 'measure-up' obesity prevention campaign in Australia. *PLoS One.* 10, 1–16 (2015)
9. Arab, U.: Geographie Information Systems (GIS) Application for Health: Case of AI Ain (UAE). *Intimai of Geoinformatics.* 7, 21–28 (2011)
10. Kumar, A., Grover, A.: Understanding the problem of diabetes in India using geo-graphical information system software. *Indian J. EndocriNol. Metab.* 18, 124–125 (2014)
11. Mining, D., Warehouse, D., Man, B.: Integrating Data Mining Techniques into Telemedicine Systems. *Inform. Econ.* 18, 120–131 (2014)
12. Youssefagha, A., Lohrmann, D., Jayawardene, W.: Use of Data Mining to Reveal Body Mass Index (BMI): Patterns Among Pennsylvania Schoolchildren, Pre-K to Grade 12. *J. Sch. Health.* 83, 85–92 (2013)
13. Su, C., Yang, C., Hsu, K., Chiu, W.: Data mining for the diagnosis of type II diabetes from three-dimensional body surface anthropometrical scanning data. *Comput. Math. with Appl.* 51, 1075–1092 (2006)