

Propuesta de una Metodología Utilizando Minería de Datos para Detectar Áreas de Oportunidad en el Proceso de Atención del Cáncer de Mama

Angélica Enriquez-Amaya¹, Raquel Torres-Peralta¹, Federico Cirett-Galán¹, Gerardo Sanchez-Schmitz¹, Jose M. Juarez²

¹Universidad de Sonora, Departamento de Ingeniería Industrial,
Rosales y Blvd. Luis Encinas S/N CP. 83000, Hermosillo, Sonora, México.
angelica.enriquez.88@gmail.com, rtorres@industrial.uson.mx,
fcirett@industrial.uson.mx, gsanchez@industrial.uson.mx

²Universidad de Murcia, Facultad de Informática.
Campus de Espinardo 30100, Murcia, España.
jmjuarez@um.es

Resumen. La minería de datos ha mejorado varios campos de la vida del ser humano basándose en patrones y tendencias descubiertas. En el cuidado de la salud, la minería de datos ayuda a la mejora de procesos reduciendo el tiempo de consulta de espera del paciente y, con la ayuda de algoritmos inteligentes, asistiendo en diagnósticos que resultan más acertados que los efectuados por un médico. En este artículo se presenta una propuesta metodológica para apoyar el proceso de atención de pacientes con cáncer de mama utilizando técnicas de minería de datos, que permita identificar áreas de oportunidad para establecer señalamientos y sugerencias de mejora en una institución de salud pública del estado de Sonora.

Palabras clave: Minería de datos, cáncer de mama, demora en la atención.

1 Introducción

Las tecnologías de información tienen una contribución importante en el cambio de paradigmas en salud pública, tanto para evitar y reducir los errores clínicos, como en la mejora de procesos y flujos de trabajo de las instituciones de salud; la minería de datos ha mejorado varios campos de la vida del ser humano, basándose en patrones y tendencias descubiertas determinando nuevos conocimientos y proporcionando diagnósticos [1]. En los países de primer mundo la implementación de minería de datos ayuda a una reducción del tiempo de espera del paciente en términos de consulta y/o diagnósticos por parte del

Enriquez-Amaya A, Torres-Peralta R, Cirett-Galán F, Sanchez-Schmitz G, Juarez JM (2018) Propuesta de una Metodología Utilizando Minería de Datos para Detectar Áreas de Oportunidad en el Proceso de Atención del Cáncer de Mama. Avances de Investigación en Ingeniería en el Estado de Sonora 4 (1):194-200

médico, lo que conlleva a una atención médica de calidad que es una de las claves del éxito de las instituciones de salud [2].

Por lo anterior, se propone desarrollar una metodología como apoyo en la atención de pacientes con cáncer de mama, que permitan identificar áreas de oportunidad para establecer señalamientos y sugerencias de mejora, utilizando técnicas de minería de datos para identificar casos prioritarios o derechohabientes con alto riesgo, así como el análisis de procesos para la optimización de recursos en el Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado de Sonora (ISSSTESON), institución de salud pública que cuenta con alrededor de 177 mil 411 derechohabientes [3].

Los apartados posteriores incluyen el fundamento teórico, una situación actual de la institución, una propuesta metodológica basada en minería de datos y finalmente se plantean los resultados y las conclusiones esperadas.

2 Marco teórico

2.1 Minería de datos

La minería de datos es el proceso de extraer conocimiento de grandes cantidades de datos, que consta de varios pasos: extracción y recolección, limpieza e integración, selección y transformación de datos, extracción de conocimiento y finalmente visualización y comunicación [4]. Los métodos de minería de datos incorporan técnicas de estadística, inteligencia artificial, aprendizaje automático (machine learning) y sistemas de bases de datos [2].

2.2 Cáncer de mama

Desde hace varias décadas la tasa de cáncer de mama ha aumentado, siendo el de mayor incidencia entre las mujeres a nivel mundial y ocupa el primer lugar entre los motivos de muerte [5], [6], [7]. Según un estudio [8], existen varios factores de riesgo biopsicosociales (factores biológicos y psicológicos [9]) en el desarrollo del cáncer de mama; los resultados del estudio determinaron que el mayor riesgo de contraer cáncer de mama fue para las mujeres de los 40-59 años. El incremento de cáncer de mama está vinculado a los estilos de vida, las transiciones demográficas y epidemiológicas; o bien a la falta de programas de detección precoz en países de bajos o medianos ingresos, donde la disponibilidad de servicios es de 35% mientras que en los países de ingresos altos es de 95% [5]. La Organización Mundial de la Salud (OMS) promueve programas nacionales para el control de cáncer de mama entre los que destaca el control integral del cáncer, que abarca la prevención, la detección precoz, el diagnóstico y tratamiento, la rehabilitación y los cuidados paliativos [10].

La demora en iniciar un tratamiento podría reducir la supervivencia en pacientes con cáncer de mama [11], [12], [13], [14]; según Ángeles-Llerenas et al. [11], el porcentaje de supervivencia para América Latina va de 66 a 78%. Según Flores-Luna et al. [14], observan que existe un 59.9% de supervivencia a cinco años, sin embargo los casos que fueron

detectados en la primera etapa (estadio clínico I) tuvieron una supervivencia mayor de 82%, mientras que los pacientes cuyo caso fue detectado en la última etapa tuvieron un 15% de supervivencia (estadio clínico IV); adicionalmente Richards et al. [15], mencionan que los pacientes que inician un tratamiento de 3 a 6 meses después de los primeros síntomas, tuvieron una supervivencia a cinco años significativamente más baja que los que esperaron menos de 3 meses.

Según Ángeles-Llerenas et al. [11], mencionan que se deben considerar 4 variables de tiempo en el retraso en la atención médica:

- tiempo total en días naturales desde el primer síntoma hasta el inicio del tratamiento;
- tiempo en días naturales desde el primer síntoma hasta la consulta con un médico acerca de los síntomas;
- el tiempo en días naturales desde la recepción de los resultados de la mamografía hasta los resultados de la biopsia diagnóstica, y
- el tiempo en días naturales desde la biopsia hasta el inicio del tratamiento.

Adicionalmente Ángeles-Llerenas et al. [11] comentan que los retrasos en la atención médica pueden ser por 2 motivos:

- retraso del paciente, que surge de un retraso en la búsqueda de atención médica después del autodescubrimiento de un posible síntoma de la enfermedad, y
- retraso del sistema, que incluye, el tiempo para obtener citas, programar pruebas de diagnóstico, recibiendo un diagnóstico definitivo e iniciar el tratamiento.

En este proyecto se pretende agilizar el proceso de atención, así como detectar derechohabientes con alto riesgo y motivarles a la autoexploración y estudios periódicos para la detección temprana de la enfermedad en la población afiliada a ISSSTESON.

2.3 Uso de la Minería de Datos

Un ejemplo práctico de la implementación de minería de datos en el ISSSTESON está en el caso de estudio de Llanes-Robles et al. [16], tiene como objetivo detectar brotes infecciosos y virales, seleccionando los casos de dengue, chinkunkunya, diarrea y rickettsiosis. El desarrollo del método para la detección de estos brotes, proveerá estrategias de medicina preventiva con base al comportamiento de la demanda de los servicios médicos en el área de consulta en los centros de atención.

Otro caso de estudio realizado también en el ISSSTESON es el de Sanz-Moreno et al. [17], cuyo objetivo es la medicina preventiva para diabetes y obesidad, la minería de datos se propone como metodología para detectar a los derechohabientes con potencial de desarrollar alguna de estas enfermedades, generando reportes periódicos para lograr una contribución valiosa a la toma de decisiones y generar estrategias de prevención.

3 Descripción del problema a abordar

El estudio se realizará en una Institución de Salud Pública del Estado de Sonora, que actualmente cuenta con bases de datos que contienen registros de los seguimientos a pacientes de cáncer de mama del año 2013 al 2018.

Aunque existe un gran número de estudios sobre la detección y tratamiento del cáncer de mama, en México no se está analizando el proceso de atención de esta enfermedad, lo que puede estar generando cuellos de botella en el tiempo de atención, desde la detección hasta que termina un tratamiento; siendo el tiempo un factor crucial para la supervivencia del paciente, que, al no medirse no puede ser mejorada. La institución desea analizar los registros disponibles con los que cuenta, con fin de tener un mayor conocimiento sobre la enfermedad entre sus derechohabientes y establecer medidas para mejorar sus procesos, de igual forma se requiere mejorar en medicina preventiva haciendo uso de las nuevas tecnologías para alcanzar una atención médica de alto nivel.

4 Propuesta de solución

En base a lo anterior se tiene como propuesta desarrollar un procedimiento apoyado en minería de datos, que permita identificar derechohabientes con alto riesgo de padecer cáncer de mama, así como áreas de oportunidad para establecer señalamientos y sugerencias de mejora en el proceso de atención de la enfermedad, a través de la exploración de los datos disponibles en los registros y seguimientos de pacientes desde el 2013.

La primera etapa consiste en analizar el comportamiento en los últimos años mediante estadística descriptiva, para tener visión del estado actual de la institución.

En la segunda etapa se prepararán los datos y se identificarán las variables que serán utilizadas por un algoritmo de agrupamiento, seleccionando a las pacientes de acuerdo a sus características y similitudes, cuáles son los aspectos que las describen y su posible relación con su estado de salud.

En la tercera etapa se identificarán las herramientas de software y algoritmos de inteligencia artificial (AI por sus siglas en inglés), se utilizarán para predecir la probabilidad de padecer la enfermedad en la población derechohabiente, y detectar a los individuos con mayor riesgo para enfocar campañas de medicina preventiva personalizadas, alentando a la revisión periódica y a una vida sana. Lo anterior lleva a elaborar informes que incluyan señalamientos y sugerencias de mejora para apoyar el proceso de atención de pacientes con cáncer de mama.

En la figura 1 se observa la metodología antes propuesta.

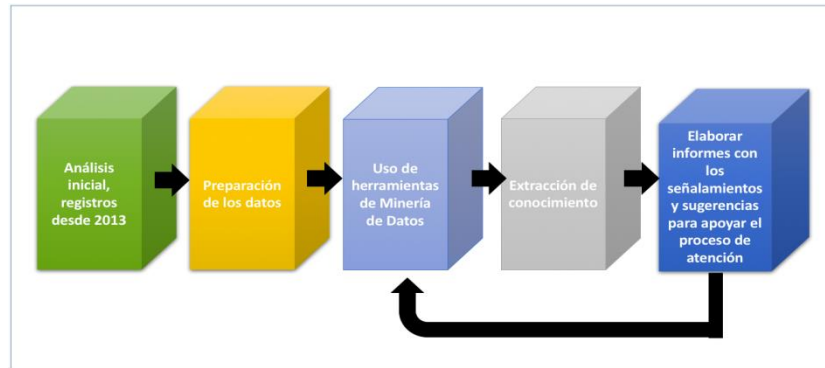


Figura 1. Propuesta metodológica para generar contenido informativo.

5 Resultados

El presente proyecto se encuentra en la fase uno de diagnóstico, y se espera que al implementar la metodología se obtengan informes con señalamientos y sugerencias de mejora, que apoyen el proceso de atención del cáncer de mama desde su detección hasta el término de su tratamiento, evaluando cada paso en la atención al paciente y detectando cuellos de botella o causas de retrasos en los procedimientos. Además, se pretende identificar a los derechohabientes con alto riesgo o casos prioritarios, promoviendo prácticas de autoexploración y chequeos periódicos para la detección precoz de la enfermedad en la población afiliada a la institución.

6 Conclusiones

Con este proyecto se espera que se obtengan algunos beneficios, como la mejora del proceso de atención del cáncer de mama, y así garantizar un rápido tratamiento, lo que podría resultar en un aumento en la probabilidad de supervivencia de las pacientes, además de ayudar a la institución a agilizar procesos para una atención pronta, y así obtener una radiografía de la cadena de atención de los pacientes con cáncer con el fin de tener una oportunidad de mejora. Por otro lado, una atención pronta puede facilitar la detección de cáncer de mama, lo que da la posibilidad de aplicar tratamientos que suelen ser más eficaces, menos complejos y menos costosos. Así mismo, una persona que fue detectada con cáncer de mama en fases tempranas y recibe un tratamiento eficaz a tiempo, puede seguir trabajando y seguir apoyando económicamente a sus familias.

El uso de los recursos tecnológicos con que contamos en la actualidad pueden contribuir considerablemente a la mejora de procesos en áreas sensibles, como la salud, ayudando a proporcionar servicios médicos de calidad.

Referencias

1. Silachan, K, y Tantasanawong, P, 2011, Evaluation Information Extraction for Health Text Categories Using C4.5 and Naïve Bayes. Proceedings Of The European Conference On Information Management & Evaluation, pp. 403-411, Business Source Complete, EBSCOhost, viewed 26 December 2017.
2. Gheorghe, M., y Petre, R., 2014. Integrating Data Mining Techniques into Telemedicine Systems. *Informatica Economica*, 18(1/2014), 120–130.
3. Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado de Sonora, 2016. Programa Institucional para la Seguridad y Servicios Sociales de los Trabajadores del Estado de Sonora.
4. Han, Jiawei., Kamber, Micheline., y Pei, Jian., 2011. *Data Mining: Concepts and Techniques*. 3rd ed. Morgan Kaufmann Publishers.
5. de la Vara-Salazar, E., Suárez-López, L., Ángeles-Llerenas, A., Torres-Mejía, G. y Lazcano-Ponce, E., 2011. Tendencias de la mortalidad por cáncer de mama en México, 1980-2009. *Salud Pública de México*, 53(5), 385–393.
6. Lorente, R. R., Hernández, M. C. R., Durán, D. H. y Lien, T. G. T., 2011. Experiencia en mujeres con cáncer de mama durante un trienio. *Revista Cubana de Obstetricia Y Ginecología*, 37(3), 349–358.
7. Guler, E. N., 2017. Gene Expression Profiling in Breast Cancer y Its Effect on Therapy Selection in Early-Stage Breast Cancer. *European Journal of Breast Health*, 13(4), 168–174.
8. Ozkan, M., Yildirim, N., Disci, R., Ilgun, A. S., Sarsenov, D., Alco, G. y Ozmen, V., 2017. Roles of Biopsychosocial Factors in the Development of Breast Cancer. *European Journal of Breast Health*, 13(4), 206–212.
9. Organización Panamericana de la Salud, 2011. Implementación del modelo biopsicosocial para la atención de personas con discapacidad a nivel nacional.
10. Organización Mundial de la Salud, nd. Cáncer de mama: prevención y control. [online]. Available from <<http://www.who.int/topics/cancer/breastcancer/es/index3.html>> [30 de octubre 2017]
11. Ángeles-Llerenas, A., Torres-Mejía, G., Lazcano-Ponce, E., Uscanga-Sánchez, S., Mainero-Ratchelous, F., Hernández-Ávila, J. E., Hernández-Ávila, M., 2016. Effect of care-delivery delay on the survival of Mexican women with breast cancer TT - Efecto de la demora en la atención sobre la supervivencia de mujeres mexicanas con cáncer de mama. *Salud Pública de México*, 58(2), 237–250.
12. Allemani, C., Weir, H. K., Carreira, H., Harewood, R., Spika, D., Wang, X.-S., Bannon, F., Ahn, J.V., Johnson, C.J., Bonaventure, A., Marcos-Gragera, R., Stiller, C., Azevedo e Silva, G., Chen, W.Q., Ogunbiyi, O.J., Rachet, B., Soeberg, M.J., You, H., Matsuda, T., Bielska-Lasota, M., Storm, H., Tucker, T.C., Coleman, M.P., the CONCORD Working Group., 2015. Global surveillance of cancer survival 1995–2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet (London, England)*, 385(9972), 977–1010.
13. McLaughlin, J.M., Anderson, R.T., Ferketich, A.K., Seiber, E.E., Balkrishnan, R., Paskett, E.D., 2012. Effect on survival of longer intervals between confirmed diagnosis and treatment initiation among low-income women with breast cancer. *J Clin Oncol* 2012;30(36):4493-4500.
14. Flores-Luna, L., Salazar-Martinez, E., Duarte-Torres, R.M., Torres-Mejía, G., Alonso-Ruiz, P., Lazcano-Ponce, E., 2008. Prognostic factors related to breast cancer survival. *Salud Publica Mex* 2008;50(2):119-125.

15. Richards, MA., Westcombe, AM., Love, SB., Littlejohns, P., Ramirez, AJ., 1999. Influence of delay on survival in patients with breast cancer: a systematic review. *Lancet*1999;353(9159):1119-1126.
16. Llanes-Robles, M.L., Torres-Peralta, R., Cirett-Galán, F., Perez-Soltero, A., 2017. Propuesta Metodológica de Análisis de Georreferenciación Apoyado con Técnicas de Minería de Datos para la Vigilancia Epidemiológica. *Avances de Investigación en Ingeniería en el Estado de Sonora*, 3(1/2017), 205-210.
17. Sanez-Moreno, E.A., Torres-Peralta, R., Cirett-Galán, F., Barceló-Valenzuela, M., 2017. Propuesta de una Metodología Utilizando Minería de Datos para Estrategias de Medicina Preventiva Más Eficaces. *Avances de Investigación en Ingeniería en el Estado de Sonora*, 3(1/2017), 211-217.