

# **Propuesta Metodológica de Análisis de Georreferenciación Apoyado con Técnicas de Minería de Datos para la Vigilancia Epidemiológica**

Mucia Lorena Llanes-Robles, Federico Miguel Cirett-Galan, Raquel Torres-Peralta,  
Alonso Perez-Soltero  
Universidad de Sonora, Departamento de Ingeniería Industrial,  
Rosales y Blvd. Luis Encinas S/N CP. 83000, Hermosillo, Sonora, México.

lorena.llanes.0@gmail.com, fcirett@industrial.uson.mx,  
rtorres@industrial.uson.mx, aperez@industrial.uson.mx

**Resumen.** La vigilancia epidemiológica es uno de los pilares de cualquier sistema de salud, en virtud de que proporciona información veraz y oportuna para la toma de decisiones adecuadas, haciendo posible la atención de la problemática específica que puede conllevar cualquier enfermedad y, la correcta evaluación del posible impacto de los programas de salud en la población. El siguiente trabajo presenta una propuesta metodológica para la detección de brotes epidemiológicos de rápida propagación, mediante el análisis de georreferenciación apoyado con técnicas de minería de datos para la mejora de toma de decisiones en medicina preventiva de una institución de salud pública del estado de Sonora.

**Palabras Claves:** Análisis de Georreferenciación, Minería de Datos, Vigilancia Epidemiológica, Medicina Preventiva, Salud Pública.

## **1 Introducción**

México se encuentra en una marcada transición poblacional y epidemiológica, la cual ha impactado de forma importante en el perfil de salud de la población [1]. La vigilancia epidemiológica es uno de los soportes de cualquier sistema de salud, proporciona la información necesaria para la toma de decisiones oportunas y adecuadas tanto para entender la problemática específica que puede implicar cualquier enfermedad como para evaluar el posible impacto de los programas de salud en la población [1].

Recientemente la vigilancia epidemiológica se convirtió en un ejercicio, técnico, científico y de vinculación que integra a múltiples fuentes de información, por medio del

Mucia Lorena Llanes-Robles, Federico Miguel Cirett-Galan, Raquel Torres-Peralta, Alonso Perez-Soltero, Propuesta metodológica de análisis de georreferenciación apoyado con técnicas de minería de datos para la vigilancia epidemiológica, 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. 205-210, 2017.

análisis, es posible estimar y predecir escenarios relevantes, que orienten las decisiones en políticas en salud [2].

Por lo anterior, se propone desarrollar una metodología de detección de brotes epidemiológicos de rápida propagación, utilizando análisis de georreferenciación apoyado con técnicas de minería de datos para la mejora de toma de decisiones.

## **2 Marco teórico**

### **2.1 Importancia de la vigilancia epidemiológica en la medicina preventiva**

Medicina Preventiva y Salud Pública se define como la ciencia y el arte de prevenir enfermedades, prolongar la vida, promover la salud y la eficiencia a través del esfuerzo comunitario organizado; se ocupa de la salud de toda la población y de la prevención de las enfermedades de las que sufre [3]. Dentro de la práctica de salud pública, es necesario que exista información oportuna sobre el curso de las enfermedades y, otros eventos de salud para poder efectuar las acciones que mejor se adapten a la necesidades, por lo que esta disciplina se encuentra dividida en varias ciencias fundamentales, entre ellas la epidemiología [4].

Kreis, et al [5] definen epidemiología como el estudio de la distribución de salud y enfermedad en poblaciones definidas, y los determinantes de esta distribución, lo que conlleva a establecer vigilancia epidemiológica como recopilar, interpretar y difundir datos que permitan la identificación temprana de amenazas potenciales para la salud pública y su impacto, al implementar dicha vigilancia, se logra la detección de “alarmas”, las cuales indican un aumento estadísticamente significativo de epidemia(s) en un área geográfica, estas alarmas proporcionan una alerta temprana de un impacto medible en la salud, permitiendo una acción eficaz de salud pública.

### **2.2 Sistemas de Información Geográfica (GIS).**

La mayoría de los datos epidemiológicos tienen una referencia de lugar y tiempo, los sistemas de información geográfica (GIS) son tecnología innovadora e ideal para manejar este tipo de información [6]. Loslier [7] define GIS como un sistema de gestión de información asistida por computadora de datos georreferenciados, en términos generales, los objetivos de un GIS son la gestión (adquisición, almacenamiento, mantenimiento), análisis (estadística, modelización espacial) y visualización (gráficos, cartografía) de datos geográficos.

El uso del análisis geográfico para estudiar la distribución de las enfermedades es esencial, los enfoques de la geografía de la salud toman una variedad de perspectivas metodológicas, algunos investigadores eligen crear modelos estadísticos de incidencias o patrones de enfermedad y contaminación ambiental, mientras que otros, adoptan enfoques cualitativos para desarrollar la comprensión de la mala salud o exploran la expresión

Propuesta metodológica de análisis de georreferenciación apoyado con técnicas de minería de datos para la vigilancia epidemiológica

geográfica de las políticas de salud (distribución, entrega, acceso a servicios de salud y población) [8].

### **2.3 Minería de Datos**

Cuando se tiene una gran cantidad de información, es necesario extraer reglas de asociación ocultas en las bases de datos para de esta forma adquirir conocimiento útil y comprensible, previamente desconocido, a este proceso de extracción se le conoce como “minería de datos”, y con los avances de sus técnicas en los últimos años, se ha comenzado a aplicar en investigaciones de la salud [9].

Las técnicas de minería de datos son particularmente apropiadas para conjuntos de grandes datos con variables que pueden interactuar de maneras complicadas. Estas técnicas buscan automáticamente relaciones que incluyan un gran número de efectos principales e interacciones complejas. Sólo se seleccionan las reglas de asociación cuyas métricas cumplen con los criterios de significación predeterminados, asegurando así que las variables que no están fuertemente correlacionadas se excluyen del modelo final [10].

## **3 Descripción del Problema**

El proyecto se desarrollará en una institución de salud pública del estado de Sonora, la Subdirección de Servicios Médicos del ISSSTESON, encargada de la medicina preventiva de la institución de salud pública, cuenta, con una gran base de datos recolectados desde hace más de una década que no ha sido considerada para la toma de decisiones, implicando campañas insatisfactorias que conllevan problemas de saturación tanto en Servicios de Urgencias como Laboratorios; además hay ausencia de la identificación temprana de amenazas potenciales, por lo que no responden de manera eficiente a las necesidades de información epidemiológica.

La Subdirección de Servicios Médicos, carece de herramientas tecnológicas que permitan un análisis eficiente de datos epidemiológicos adaptado a las necesidades específicas de dicha institución de salud pública, que permita contar con información confiable y oportuna para el control eficaz de brotes infecciosos y virales, a la vez, no se cuenta con una detección de las zonas geográficas que son más propensas a epidemias.

## **4 Propuesta de Solución**

En base a lo previamente mencionado, este proyecto contempla desarrollar un método para la detección de brotes infecciosos y virales, mediante la tamización de grandes cantidades de datos siguiendo los pasos mostrados en la sección 4.4.

El estudio se lleva a cabo con los datos provenientes del sistema de información de consultas de ISSSTESON de los pacientes del municipio de Hermosillo localizada en el estado de Sonora, México; se encuentra en las coordenadas geográficas 29.08919,-

110.96133. El municipio de Hermosillo representa el 8.72% de la superficie del estado, colinda al norte con los municipios de Pitiquito, Carbó y San Miguel de Horcasitas; al este con los municipios de San Miguel de Horcasitas, Ures, Mazatán, La Colorada y Guaymas; al sur con el de Guaymas y el Golfo de California y al oeste con el Golfo de California y el municipio de Pitiquito. El municipio de Hermosillo cuenta con 5 localidades principales las cuales son Hermosillo, Miguel Alemán, Bahía de Kino, San Pedro el Saucito y Ejido La Victoria.

#### **4.1 Adquisición de Datos**

Los datos sobre las incidencias epidemiológicas y datos demográficos de los años 2013-2017, se adquirieron del departamento de estadística dentro de la Subdirección de Servicios Médicos del instituto de salud pública.

#### **4.2 Preparación y Manipulación de Datos**

En la fase actual de la investigación, se han seleccionado, primeramente, los casos de Dengue, Chikungunya, Diarrea y Rickettsiosis registrados entre 2013 y 2017 para su análisis. Además del tipo de epidemia, se cuenta con otros conjuntos de datos, como grupo de edad, sexo, fecha de contagio, colonia/zona proveniente, por mencionar algunos.

#### **4.3 Creación de mapas de distribución epidemiológica basados en GIS**

Un GIS considera las diferentes maneras en que los conjuntos de datos pueden ser asociados a cierta ubicación, en el presente estudio, los datos epidemiológicos serán vinculados a colonias y/o zonas.

Primero se determinarán los casos epidemiológicos y se organizarán las bases de datos, posteriormente, los datos epidemiológicos se transferirán a un mapa base mediante la ayuda de un software, en este mapa, habrá información geográfica que representen los límites y colonias/zonas.

#### **4.4 Metodología**

La metodología propuesta se basa en el método descrito en la investigación de Buczak et al. [10,11], a continuación, se hace una descripción de cada uno de los pasos (Ver Figura 1).

1. Identificar variables relevantes para su posterior análisis: Se realiza una revisión de la literatura para identificar las variables que están correlacionadas con la incidencia de la epidemia.

2. Construcción del modelo: Parte principal donde se encuentran los elementos de georreferenciación y minería de datos. Los datos son pre-procesados y utilizados para encontrar reglas de asociación, un subconjunto de estas reglas bajo enfoques orientados al mejoramiento de las actuales estrategias de medicina preventiva se selecciona para crear un clasificador, que posteriormente se convierte en el modelo de detección.
  - a) Preprocesamiento de datos: Los datos se procesan previamente para convertirlos en la resolución espaciotemporal deseada.
  - b) Extracción de reglas: Aplicación de técnicas de minería de datos para extraer automáticamente reglas de asociación de los datos.
  - c) Clasificador: Para la construcción del clasificador, primeramente, es necesario determinar que reglas, de las extraídas anteriormente, deben ser utilizadas y en que secuencia usarlas.
3. Generación de detecciones: Los clasificadores finales que utilizan las reglas seleccionadas descritas en el paso anterior, se convierten en los modelos que generan detecciones de brotes epidemiológicos. Estos modelos son los que serán evaluados bajo métricas establecidas.
4. Detectar, de forma periódica, las zonas geográficas que presenten una mayor incidencia de casos con el fin de orientar campañas de medicina preventiva y acciones de contención de epidemias de rápida propagación.

Con el desarrollo de un método para la detección de brotes infecciosos y virales se proveerán estrategias de medicina preventiva en base al comportamiento de la demanda de los servicios médicos en el área de consulta en los centros de atención.

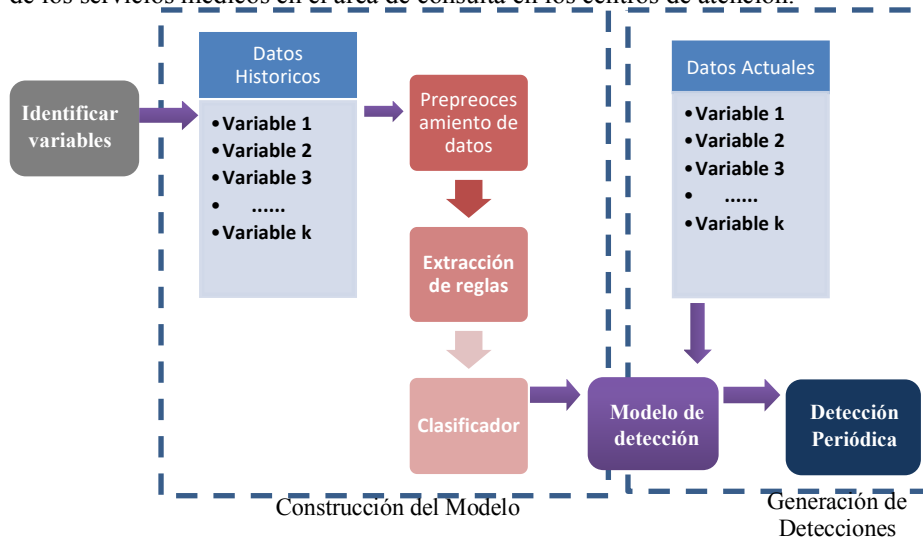


Figura 1. Etapas de la metodología propuesta.

## 5 Resultados y Beneficios Esperados

El presente proyecto forma parte de uno con mayor alcance y constituido por un enfoque multidisciplinario, se espera que, además de complementar una parte fundamental del proyecto mayor, se logre la emisión de alertas y preparación de una pronta y eficaz respuesta, orientar y fundamentar las acciones de salud a través de la planeación y evaluación de los diferentes problemas mediante la investigación, buscando beneficiar el desarrollo, el bienestar social y la calidad de vida de la comunidad en general.

## 6 Conclusiones

Los datos sobre la incidencia de epidemias pasadas, junto con las variables ambientales y socioeconómicas, que se han demostrado en la literatura para influir en la exposición (dispersión, propagación) potencial a una epidemia, pueden proporcionar suficiente información para permitir una predicción, de si es probable que exista una alta incidencia de cierta epidemia en un determinado tiempo y lugar en el futuro.

La aplicación de minería de datos, apoyada con, análisis de georreferenciación en las incidencias de epidemias prevalentes en la base de datos recolectada por la Subdirección de Servicios Médicos, permitirá el desarrollo de un sistema de vigilancia epidemiológica y el favorecimiento de las estrategias de medicina preventiva.

### Referencias

1. Kuri Morales P. La vigilancia epidemiológica en México: elemento central para enfrentar problemas globales de salud (2006).
2. Secretaria de Salud. Programa de Acción Específico Vacunación Universal. 2014. p. 72.
3. Gidey G, Taju S, Hagos AS. Introduction to Public Health (2005)
4. Beaglehole R, Bonita R, Kjellstrom T. Basic epidemiology, 2nd ed. World Heal. Organ. (2006)
5. Kreis IA, Araceli B, Leonardi GS, Meara J, Murray V. Essentials of Environmental Epidemiology for Health Protection. Oxford University Press (2013).
6. Nobre FF, Carvalho MS. Spatial and Temporal Analysis of Epidemiological Data. 2016;5–10.
7. Loslier L. Geographical Information Systems ( GIS ) from a Health Perspective. 2016;5–10.
8. Yagoub MM. Geographie Information Systems ( GIS ) Application for Health : Case of AI Ain ( UAE ). Int. J. Geoinformatics. 2011;7:21–8.
9. Huang YC. The application of data mining to explore association rules between metabolic syndrome and lifestyles. Heal. Inf. Manag. J. 2013;42:29–36.
10. Buczak AL, Baugher B, Babin SM, Ramac-Thomas LC, Guven E, Elbert Y, et al. Prediction of High Incidence of Dengue in the Philippines. PLoS Negl. Trop. Dis. 2014;8.
11. Buczak AL, Baugher B, Guven E, Ramac-Thomas LC, Elbert Y, Babin SM, et al. Fuzzy association rule mining and classification for the prediction of malaria in South Korea. BMC Med. Inform. Decis. Mak. 2015;15:47.