

UNIVERSIDAD DE SONORA DIVISIÓN DE INGENIERÍA



POSGRADO EN INGENIERÍA INDUSTRIAL MAESTRÍA EN INGENIERÍA EN SISTEMAS Y TECNOLOGÍA

UNA METODOLOGÍA PARA LA DETECCIÓN DE BROTES
EPIDEMIOLÓGICOS DE RÁPIDA PROPAGACIÓN UTILIZANDO
MINERÍA DE DATOS Y ANÁLISIS DE GEORREFERENCIACIÓN

TESIS

PRESENTADA POR

MUCIA LORENA LLANES ROBLES

Desarrollada para cumplir con uno de los
requerimientos parciales para obtener
el grado de Maestra en Ingeniería

DIRECTOR DE TESIS
DR. FEDERICO CIRETT GALÁN

CODIRECTOR DE TESIS
DR. ALONSO PÉREZ SOLTERO

HERMOSILLO, SONORA, MÉXICO.

NOVIEMBRE 2018

Repositorio Institucional UNISON



**"El saber de mis hijos
hará mi grandeza"**



Excepto si se señala otra cosa, la licencia del ítem se describe como openAccess



Hermosillo, Sonora a 4 de octubre de 2018

MUCIA LORENA LLANES ROBLES

Con fundamento en el artículo 66, fracción III, del Reglamento de Estudios de Posgrado vigente, otorgamos a usted nuestra aprobación de la fase escrita del examen de grado, como requisito parcial para la obtención del Grado de Maestra en Ingeniería.

Por tal motivo este jurado extiende su autorización para que se proceda a la impresión final del documento de tesis: **UNA METODOLOGÍA PARA LA DETECCIÓN DE BROTES EPIDEMIOLÓGICOS DE RÁPIDA PROPAGACIÓN UTILIZANDO MINERÍA DE DATOS Y ANÁLISIS DE GEORREFERENCIACIÓN** y posteriormente efectuar la fase oral del examen de grado.

ATENTAMENTE

Dr. Federico Miguel Cirett Galán
Director de Tesis y Presidente del Jurado

Dr. Alonso Pérez Soltero
Codirector y Vocal del Jurado

Dra. Raquel Torres Peralta
Secretaría del Jurado

Dr. Mario Barceló Valenzuela
Vocal del Jurado



Barcelona, España, a 29 de junio de 2018.

MUCIA LORENA LLANES ROBLES

Con fundamento en el artículo 66, fracción III, del Reglamento de Estudios de Posgrado de la Universidad de Sonora, otorgo a usted mi aprobación de la fase escrita del examen profesional, como requisito parcial para la obtención del Grado de Maestra en Ingeniería.

Por tal motivo, como sinodal externo y vocal del jurado, extiendo mi autorización para que se proceda a la impresión final del documento de tesis: **UNA METODOLOGÍA PARA LA DETECCIÓN DE BROTES EPIDEMIOLÓGICOS DE RÁPIDA PROPAGACIÓN UTILIZANDO MINERÍA DE DATOS Y ANÁLISIS DE GEORREFERENCIACIÓN** y posteriormente efectuar la fase oral del examen de grado.

ATENTAMENTE,



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament de Ciències de la Computació

Dr. MIQUEL SÀNCHEZ-MARRÈ FIEMSS
UNIVERSITAT POLITÈCNICA DE CATALUNYA · BARCELONATECH
Sinodal Externo y Vocal del Jurado

Edifici OMEGA, Campus Nord
08034 Barcelona
E-08034 Barcelona

RESUMEN

Dentro del modelo de atención a la salud en México encontramos la prevención y la promoción de la salud como dos componentes esenciales, un elemento sustancial de la prevención de la salud es su carácter anticipatorio, el cual busca atender, no a la epidemia directamente, sino a los determinantes de la salud, para fortalecer los que le son favorables y limitar o eliminar aquellos que no lo son.

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.

La problemática consiste en la carencia de un análisis eficiente de datos epidemiológicos que permita contar con información confiable y oportuna para el control eficaz de brotes infecciosos y virales, a la vez, de la detección de las zonas geográficas que son más propensas a epidemias.

El presente documento muestra el desarrollo y aplicación de una metodología adaptada, la cual se basó principalmente en los pasos o etapas del proceso llamado Knowledge Discovery in Databases (KDD), implementadas según se indica en la metodología de Buczak *et al.* (2012; 2014; 2015), diseñada para el desarrollo de procesos de Minería de datos. Primeramente, se realizó una selección de las fuentes de datos para posteriormente hacer la limpieza y preprocesamiento de ellos, y así lograr dejarlos en el formato adecuado para poder aplicar los algoritmos de Minería de Datos y lograr la detección de brotes epidemiológicos de rápida propagación.

Al aplicar con éxito las técnicas de minería de datos, se obtuvo como resultado la identificación de los pacientes con características similares generando conocimiento útil a través de la caracterización de las epidemias de rápida propagación.

ABSTRACT

Within the health care model in Mexico we find the prevention and promotion of health as two essential components, a substantial element of health prevention is its anticipatory character, which seeks to address, not the epidemic directly, but to the determinants of health, to strengthen those who are favorable and limit or eliminate those who are not.

Epidemiological surveillance is one of the supports of any health system, it provides the necessary information for the timely and appropriate decision making to understand the specific problems that any disease can imply and to assess the potential impact of health programs on the population.

The problem is the lack of an efficient analysis of epidemiological data that allows to have reliable and timely information for the effective control of infectious and viral outbreaks, at the same time, of the detection of the geographical zones that are more Prone to epidemics.

This document shows the development and application of an adapted methodology, which was based primarily on the steps or stages of the process called Knowledge Discovery in Databases (KDD), implemented as indicated in the Buczak et al methodology. (2012; 2014; 2015), designed for the development of data mining processes.

First, a selection of the data sources was made for later cleaning and preprocessing of them, and thus to be able to leave them in the appropriate format in order to apply the data mining algorithms and to achieve the detection of outbreaks Fast-spreading epidemiological.

By successfully applying the techniques of data mining, it was obtained as a result the identification of patients with similar characteristics generating useful knowledge through the characterization of epidemics of rapid spread.

Dedicada a

A Dios y mi familia,

en especial a Kyra, mi princesa, mi bebe, este trabajo es fruto tuyo.

AGRADECIMIENTOS

Agradezco profundamente a mi madre, hermano y José Antonio quienes me demostraron su apoyo, cariño y ejemplo.

A ti Michelle, por tu cariño y comprensión, sin mencionar el apoyo incondicional que siempre me has demostrado, la fe que siempre tuviste en mí, fue uno de los principales factores que me ayudo en salir adelante, gracias. A mis amigos, gracias por todos los momentos gratos a su lado en estos años, en especial a ti, mi Mayra, quien siempre me ha apoyado en todo, siempre tan pendiente de mí, aunque se haya vuelto sumamente difícil reunirnos como antes lo hacíamos, sé que tu amistad nunca cambiará. Y, por último, a ti Katia Alejandra por estar apoyándome en los momentos finales de esta maravillosa etapa.

Al Dr. Federico Cirett, por ser un gran director de tesis, por su disposición, sus consejos, su amabilidad y apoyo permanente, haciendo posible el desarrollo y culminación de este trabajo, y a la Dra. Raquel Torres, por su profesionalismo, sus comentarios tan valiosos, su asesoría en algunos temas de la tesis, y por ser una gran mentora académica.

Al Dr. Mario Barceló, quien me guio en la estructura del documento, sin importar todas las veces que sufrí de una crisis existencial como escritora, encontró la forma de indicarme el camino correcto, siempre le estaré agradecida por todas sus enseñanzas, su tiempo y sus observaciones tan precisas.

A todo el personal de la Maestría en Ingeniería en Sistemas y Tecnologías, por su apoyo académico, sus enseñanzas, al coordinador Dr. Alonso Pérez Soltero, por estar tan pendiente de nosotros, facilitarnos los tramites y su apoyo incondicional al siempre buscar la manera de ayudarnos. Al Dr. Miquel Sànchez-Marrè por ser un gran tutor, sus excelentes consejos, orientación y recibimiento durante la estancia en Barcelona.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) y al Programa de Fortalecimiento de la Calidad Educativa (PFCE) por su apoyo económico.

ÍNDICE GENERAL

RESUMEN	ii
ABSTRACT	iii
AGRADECIMIENTOS	v
ÍNDICE GENERAL	vi
ÍNDICE DE FIGURAS	viii
ÍNDICE DE TABLAS	xi
2. INTRODUCCIÓN	1
2.1. Presentación	1
2.2. Planteamiento del problema.....	2
2.3. Objetivo general	3
2.4. Objetivos específicos	3
2.5. Hipótesis	4
2.6. Alcances y delimitaciones	4
2.7. Justificación.....	5
3. MARCO DE REFERENCIA	6
3.1. Epidemiología y Vigilancia Epidemiológica en el ámbito de la Salud Pública. 6	
3.2. Epidemiología de las enfermedades de rápida propagación.....	8
2.2.1. Algunas de las Principales enfermedades de rápida propagación	8
3.3. Sistemas de Información Geográfica (GIS).....	15
3.3.1. Sistemas de Información Geográfica (GIS) como herramienta dentro de la Salud Pública.....	16
3.4. Minería de Datos.....	19
3.4.1. Definición	19
3.4.2. Proceso de Descubrimiento de Conocimiento en Base de Datos.	20
3.4.3. Modelos y tareas de minería de datos.....	21
3.4.4. Técnicas de minería de datos.....	25

3.4.5. Minería de Datos como herramienta dentro de la Salud Pública.	26
3.5. Análisis de Datos	27
3.5.1. Estadística Descriptiva	27
3.6. Software	28
3.7. Análisis de Estudios Previos	31
4. METODOLOGÍA	32
4.1. Fase I. Identificar variables relevantes para su posterior análisis.	34
4.2. Fase II. Construcción del modelo	35
4.3. Fase III. Generación de detecciones	36
5. IMPLEMENTACIÓN	39
5.1. Fase I. Identificar variables relevantes para su posterior análisis.	39
5.2. Fase II. Construcción del modelo	41
Paso 1. Preprocesamiento de datos	42
Paso 2. Aplicación de técnicas de minería de datos	61
6. CONCLUSIONES, RECOMENDACIONES Y TRABAJOS FUTUROS	68
6.1. Conclusiones	68
6.2. Recomendaciones	69
6.3. Trabajos futuros	71
7. REFERENCIAS	72
8. ANEXOS	79
8.1. Anexo A	79
8.2. Anexo B	81
8.3. Anexo C	93

ÍNDICE DE FIGURAS

Figura 1.1. Planteamiento del Problema	3
Figura 2.1. Mosquito Aedes Aegypti, vector que transmite dengue, chikungunya y zika (Dirección General de Promoción de la Salud, 2016)	8
Figura 2.2. Cronología de los síntomas e infección (Weaver y Lecuit 2016).	11
Figura 2.3. Períodos de incubación extrínseco e intrínseco (Pan American Health Organization, 2011).....	12
Figura 2.4. Ejemplo de una representación espacio-temporal (Zhou et al., 2015) ...	18
Figura 2.5. Ejemplo de representación espacio-temporal (Liu et al., 2015).....	18
Figura 2.6. Etapas para el proceso de extracción del conocimiento (Fayyad, Piatetsky-Shapiro y Smyth,1996).....	20
Figura 2.7. Modelos de Conocimiento (Gibert, Sànchez-Marrè y Codina, 2006).....	21
Figura 2.8. Representación de la Regresión lineal simple (Fayyad, Piatetsky-shapiro y Smyth, 1996).....	23
Figura 2.9. Representación de la Clasificación lineal simple (Fayyad, Piatetsky-shapiro y Smyth, 1996)	23
Figura 2.10. Distintas formas de representar clúster (Witten y Frank, 2005).....	25
Figura 3.1. Etapas de la metodología propuesta, adaptado de Buczak et al. (2012; 2014; 2015).....	34
Figura 3.2. Desarrollo del paso dos de la metodología	36
Figura 3.3. Desarrollo de pasos en la generación de detecciones	36
Figura 3.4. Detección periódica	37
Figura 4.1. Estaciones Meteorológicas en Sonora registradas en CLICOM (Datos del mapa: Google, INEGI).....	40
Figura 4.2 Estaciones Meteorológicas Automáticas (EMA).....	41
Figura 4.3 Variación mensual de la precipitación entre los años 2014 y 2017 en la ciudad de Hermosillo, Sonora, México. Fuente: Base de Datos Climatológica Nacional (CLICOM).....	43

Figura 4.4. Variación estacional de la temperatura promedio mensual entre los años 2014 y 2017 en la ciudad de Hermosillo, Sonora, México. Fuente: Base de Datos Climatológica Nacional (CLICOM).....	43
Figura 4.5. Variación estacional de la temperatura máxima mensual entre los años 2014 y 2017 en la ciudad de Hermosillo, Sonora, México. Fuente: Base de Datos Climatológica Nacional (CLICOM).....	44
Figura 4.6. Variación estacional de la temperatura mínima mensual entre los años 2014 y 2017 en la ciudad de Hermosillo, Sonora, México. Fuente: Base de Datos Climatológica Nacional (CLICOM).....	45
Figura 4.7. Comportamiento del número total de casos de dengue registrados entre los años 2014 y 2017 en el municipio de Hermosillo.....	46
Figura 4.8. Capas superpuestas de las colonias que presentaron diagnósticos durante el 2014, 2015, 2016, 2017 y 2018	47
Figura 4.9 Intersección geométrica de las colonias con diagnósticos en los años 2014, 2015, 2016, 2017 y 2018.....	48
Figura 4.10 Mapa de los casos diagnosticados con Chikungunya entre los años 2014 y 2017	49
Figura 4.11 Mapa de los casos diagnosticados con Rickettsiosis entre los años 2014 y 2017	50
Figura 4.12. Comparación entre los años 2016 y 2017 con respecto a la temperatura	51
Figura 4.13. Comportamiento mensual de la rinofaringitis entre los años 2014 y 2017 en el municipio de Hermosillo, Sonora	52
Figura 4.14. Casos de rinofaringitis por grupo de edad registrados entre los años 2014 y 2018 en el municipio de Hermosillo, Sonora	53
Figura 4.15. Casos de rinofaringitis por tipo de derechohabiente registrados entre los años 2014 y 2018 en el municipio de Hermosillo, Sonora.....	53
Figura 4.16. Registro de diagnóstico mensual durante el 2017 de rinofaringitis en trabajadores de las seis dependencias más significativas	53

Figura 4.17. Diagnostico mensual de rinofaringitis en trabajadores del H. Ayuntamiento durante el año 2017	54
Figura 4.18. Diagnostico mensual de rinofaringitis en trabajadores del Gobierno de Sonora durante el año 2017	54
Figura 4.19. Colonias que registraron diagnósticos de rinofaringitis durante el año 2017 en la ciudad de Hermosillo, Sonora.....	55
Figura 4.20. Colonias con mayor incidencia de rinofaringitis durante el 2017 en la ciudad de Hermosillo, Sonora	56
Figura 4.21. Colonias con mayor incidencia durante el mes de enero del 2017 en la ciudad de Hermosillo, Sonora	57
Figura 4.22. Colonias con mayor incidencia durante el mes de diciembre del 2017 en la ciudad de Hermosillo, Sonora.....	57
Figura 4.23. Colonias con mayor cantidad de diagnósticos tanto en enero como en diciembre del 2017 en la ciudad de Hermosillo, Sonora	58
Figura 4.24. Resultados obtenidos de Comorbilidad con mayor significancia entre la Rinofaringitis, Faringitis Aguda, Diarrea, Dengue, Rickettsiosis, Chikungunya y Obesidad.....	59
Figura 4.25. Mapa de Calor según la comorbilidad encontrada entre las distintas enfermedades	59
Figura 4.26. Resultado del clúster con k=3	62
Figura 4.27. Dendograma obtenido en RStudio	63
Figura 4.28. Dendograma obtenido en SPSS.....	64
Figura 4.29. Resultados confiabilidad de los árboles de decisión	65
Figura 4.30. Árbol de Decisión obtenido en el software Weka	66
Figura 4.31. Árbol de Decisión Obtenido en el Software RapidMiner	67
Figura 7.1. Plantilla Informe (Tema Intermedio) versión original	79

ÍNDICE DE TABLAS

Tabla 2.1 Matriz comparativa para las definiciones de Minería de Datos	20
Tabla 2.2. Relación existente entre las tareas y técnicas de minería de datos (Adaptado de Hernández, Ramírez y Ferri, 2004).....	26
Tabla 2.3. Matriz de Software	30
Tabla 4.1. Fuente de Datos.....	39
Tabla 4.2. Derechohabientes que presentaron recurrencia en diagnósticos de rinofaringitis durante el 2017 en el municipio de Hermosillo, Sonora	54
Tabla 4.3. Colonias con mayor cantidad de diagnósticos de rinofaringitis en enero y diciembre del año 2017	57
Tabla 4.4. Derechohabientes de tipo trabajador con mayor recurrencia de diagnóstico	58
Tabla 4.5. Caracterización por enfermedad	60

1. INTRODUCCIÓN

En este primer capítulo se desarrollan los puntos iniciales del proyecto de tesis. Inicialmente se desarrolla la ambientación de la realidad del problema, en relación con el medio dentro del cual aparece, lo cual da pie al siguiente punto donde se expone el problema que se pretende resolver utilizando Sistemas de Información Geográfica (GIS) apoyado con técnicas de minería de datos. Posteriormente se presentan los objetivos, general y específicos. Después se menciona la hipótesis seguido de los alcances y delimitaciones de la investigación. En última instancia se plantea la justificación del proyecto.

1.1. Presentación

El proyecto de investigación se desarrolló en una institución de salud pública del estado de Sonora, el cual es un organismo público descentralizado con personalidad jurídica y patrimonio propio, tiene a su cargo el otorgamiento de las prestaciones y servicios de seguridad social, mayormente pensiones y servicios de salud, en favor de los servidores públicos del Gobierno del Estado de Sonora y sus respectivos familiares, Ayuntamientos y Organismos afiliados, actualmente proporciona atención médica a 177,201 derechohabientes al mes de marzo de 2017.

La Subdirección de Servicios Médicos, encargada de la medicina preventiva de la institución de salud pública, ha recopilado datos de los derechohabientes desde hace más de una década por medio del sistema de consultas médicas, llegando a conformar una gran base de datos, la cual no han aprovechado para la extracción de información que ayude en la toma de decisiones.

A finales del 2015 la Secretaria de Salud del Estado de Sonora emitió la alerta de Dengue y Chikungunya, la institución de salud pública donde se lleva a cabo este proyecto, otorgó mil 924 consultas en el Servicio de Urgencias con diagnóstico probable de alguna de estas epidemias durante ese año; en promedio se otorgaron 150 consultas al mes; en septiembre se registraron 402 y en los primero seis días de

octubre se convirtieron en 313, ocasionando problemas de saturación tanto en Servicios de Urgencias como Laboratorios en la ciudad de Hermosillo (ISSSTESON, 2015).

El anterior panorama descrito (ISSSTESON, 2015) hizo evidente la carencia de una manera eficiente de analizar los datos epidemiológicos, que permita contar con información confiable y oportuna para el control eficaz de brotes infecciosos y virales, como faringitis, dengue y chikungunya, por mencionar algunas, a la vez, no se cuenta con una detección de las zonas geográficas que son más propensas a enfermedades, por lo que no se realizan campañas satisfactorias, tomando en cuenta lo descrito, no se puede tener una identificación temprana de amenazas potenciales para la salud pública. Las repercusiones de estos problemas se hacen presentes en la saturación de los servicios que esta institución brinda, ocasionando dificultad para manejar los casos que se presentan.

La falta de identificación y aislamiento de nuevos casos de enfermedades de rápida propagación, contribuyen a una respuesta poco efectiva ante diversos problemas de salud pública, no se responde de manera eficiente a las nuevas necesidades de información epidemiológica, impidiendo satisfacer las nuevas exigencias de datos para la extracción de información utilizable en escenarios futuros.

1.2. Planteamiento del problema

En la figura 1.1 se observa que la Subdirección de Servicios Médicos cuenta con una base de datos necesaria para la obtención de información epidemiológica, pero carece de una metodología para analizarla y les permita contar con la detección eficaz de brotes infecciosos y virales.

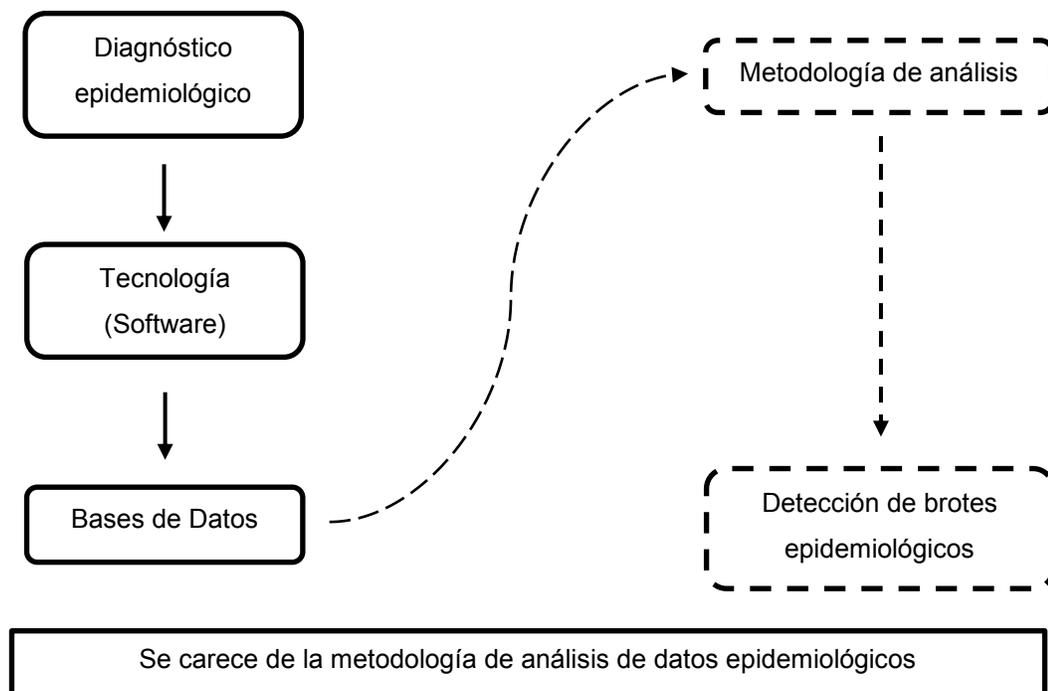


Figura 1.1. Planteamiento del Problema

1.3. Objetivo general

Desarrollar una metodología de detección de brotes epidemiológicos de rápida propagación, utilizando análisis de georreferenciación y minería de datos para la mejora de toma de decisiones en la Subdirección de Servicios Médicos.

1.4. Objetivos específicos

- Establecer procedimientos para el análisis de datos y detección de un comportamiento atípico en la demanda de atención.
- 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 enfermedades de rápida propagación.
- Desarrollar una metodología que sea objetiva y reproducible, para proporcionar estrategias de medicina preventiva que puedan ser implementadas en otras instituciones de salud.

- Validar la metodología resultante mediante la utilización de métricas adecuadas según el enfoque de investigación.

1.5. Hipótesis

El uso del análisis de georreferenciación, apoyado con técnicas de minería de datos, en las incidencias de enfermedades prevalentes, permitirá el desarrollo de un sistema de vigilancia epidemiológica y el favorecimiento de las estrategias de medicina preventiva.

1.6. Alcances y delimitaciones

En este proyecto se contempla desarrollar un método para la detección de brotes infecciosos y virales, como rinofaringitis, dengue, zika, chikungunya y cualquier enfermedad que surja (apegada a la duración de la investigación) y la institución considere necesario vigilar, mediante la tamización de grandes cantidades de datos.

Algunas limitaciones del proyecto son:

- La información recabada proveniente de la institución es exclusivamente de la Subdirección de Servicios Médicos; los datos registrados son de trabajadores del Estado y sus familiares directos, por lo que no es una muestra de toda la población de Sonora.
- Existe información que no se podrá utilizar o revelar con el propósito de cuidar la confidencialidad de los derechohabientes.
- Los resultados de este proyecto de tesis son exclusivamente para ayudar en la toma de decisiones en medicina preventiva.
- Acceso a información disponible en las bases de datos de la Universidad de Sonora y aquellas libres.

1.7. Justificación

El presente proyecto forma parte de uno con mayor alcance y constituido por un enfoque multidisciplinario, esta parte se desarrollará para detectar brotes de enfermedades infecciosas y virales, en determinados sectores geográficos, así como evaluar su propagación, esto, con el fin de emitir alertas y preparar 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.

2. MARCO DE REFERENCIA

En este apartado se recopila y expone la base teórica sobre la problemática, misma que sustentará la metodología que será aplicada. En un principio se presenta la información sobre el desenvolvimiento de la vigilancia epidemiológica en el ámbito de la salud pública, seguida de la epidemiología de las enfermedades que serán analizadas para la detección de sus brotes. Seguido de los conceptos de minería de datos, sus principales fundamentos, se revisan modelos, tareas y técnicas más comúnmente aplicados. Asimismo, se describen los GIS y su uso como herramienta dentro de la salud pública.

2.1. Epidemiología y Vigilancia Epidemiológica en el ámbito de la Salud Pública.

Para poder hablar del concepto de epidemiología es necesario establecer el termino de *Medicina Preventiva y Salud Pública*, el cual 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 (Gidey, Taju y Hagos, 2005). 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* (Bonita, Beaglehole y Kjellström, 2006).

Kreis, et al. (2013) 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 activación 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.

Dentro de dicho ámbito, la epidemiología se centra en el control de enfermedades y la ejecución y evaluación de programas (Khoury *et al.*, 2013), las instituciones de salud pública la utilizan de manera regular con el propósito de identificar la salud en localidades y poder solucionar problemas, para ello, es necesario recopilar y analizar datos detallados para determinar si los servicios de salud se encuentran disponibles, accesibles, eficaces y eficientes. La esencia de la vigilancia de la salud pública es el uso de datos para monitorear los problemas de salud y, de esta manera, facilitar su prevención o control. Los datos y las interpretaciones derivadas de la evaluación de los datos de vigilancia pueden ser útiles para establecer prioridades, planificar y llevar a cabo programas de control de enfermedades y evaluar la eficacia de los esfuerzos de control (Centers for Disease Control and Prevention, 2006).

En México, la vigilancia epidemiológica es entendida como una actividad indispensable para cualquier servicio de salud, por ser la proveedora de información para la toma de decisiones. Se lleva a cabo a través del Sistema Nacional de Vigilancia Epidemiológica (SINAVE), sistema de inteligencia capaz de proporcionar información sobre morbilidad y mortalidad, pero dicho sistema ha enfrentado grandes retos a través de los años, que van desde el recurso humano que lo opera hasta la incorporación y modernización de infraestructura informática y procesos de recopilación, administración y difusión de información (Kuri Morales, 2006).

La epidemiología ha evolucionado en México debido a la existencia simultánea de enfermedades por causas infecciosas y parasitaria, como enfermedades dentro del grupo crónico-degenerativas, los últimos retos planteados para la vigilancia epidemiológica en este país son el control y prevención de las enfermedades transmisibles y la disminución en el número de las enfermedades no transmisibles, (Sistema Nacional de Vigilancia Epidemiológica, 2014).

2.2. Epidemiología de las enfermedades de rápida propagación.

Según Bonita, Beaglehole y Kjellström (2006) las enfermedades infecciosas son causadas por la transmisión de un agente patógeno específico a un huésped susceptible. Los seres humanos pueden adquirir estos agentes infecciosos de dos maneras; directamente, de otros humanos o animales infectados, e indirectamente, a través de vectores (organismos vivos), vehículos o partículas en el aire.

El desplazamiento de los seres humanos facilita y favorece la rápida propagación de enfermedades entre distintos puntos del planeta, a continuación, se presentan algunas de las principales enfermedades de rápida propagación.

2.2.1. Algunas de las Principales enfermedades de rápida propagación

Las enfermedades transmitidas por vector (ETV) comprenden un grupo variado de enfermedades que difieren en sintomatología y patología, en las que están involucrados una extensa variedad de agentes etiológicos y transmisores (Dirección General de Epidemiología, 2016).

Los vectores son organismos vivos que sirven para transmitir un patógeno de un huésped a un ser humano, animal o ambos (CropLife International, 2016); un ejemplo de vector es el mosquito *Aedes Aegypti* (Figura 2.1.).



Figura 2.1. Mosquito *Aedes Aegypti*, vector que trasmite dengue, chikungunya y zika (Dirección General de Promoción de la Salud, 2016)

Mientras que los vehículos, son objetos contaminados o elementos del medio ambiente que sirven para transmitir la enfermedad, como agua, alimentos, ropa, sangre, plasma, instrumentos quirúrgicos, entre otros (Centers for Disease Control and Prevention, 2006).

La información médica descrita para cada una de las enfermedades presentadas a continuación se exhibe mediante un formato estandarizado el cual se tomó del libro *Control of communicable diseases manual* (Heymann, 2008), mostrando la siguiente información para cada enfermedad:

Nombre de la enfermedad.

1. Descripción. Principales signos clínicos de la enfermedad, diferenciándola de otras que pudieran tener un cuadro similar.

2. Distribución. Datos sobre los lugares en que prevalece la enfermedad y los grupos de población a los que ataca con mayor frecuencia.

3. Reservorio. Indica el ser humano, animal, artrópodo, planta, tierra, sustancia o una combinación de todos ellos, finales, intermedios o de ambos tipos, que constituyen el punto de partida de la infección para el huésped susceptible.

4. Modo de transmisión. Mecanismos por los cuales el agente infeccioso se extiende y llega a los seres humanos. Estos mecanismos son de tipo directo o indirecto.

5. Período de incubación. Intervalo de incubación en horas, días o semanas entre la exposición inicial y efectiva al agente infeccioso y el primer signo o síntoma de la infección.

6. Período de transmisibilidad. Lapso en días, semanas o meses en el cual se transmite el agente infeccioso, de manera directa o indirecta, de una persona infectada a otra, de un animal infectado a los seres humanos, o de una persona infectada a los animales o a los artrópodos.

- **Dengue.**

1. Descripción. Enfermedad vírica febril y aguda que se caracteriza por comienzo repentino, fiebre que dura de tres a cinco días (rara vez más de siete), cefalea intensa, mialgias, artralgias, dolor retro-orbital, anorexia, alteraciones del aparato gastrointestinal y erupción. Las epidemias tienen carácter “explosivo”, pero la tasa de letalidad es muy baja siempre que no se presente dengue hemorrágico. La susceptibilidad parece ser universal en los seres humanos, pero los niños suelen tener una enfermedad más benigna que los adultos. El restablecimiento de la infección por un serotipo genera inmunidad homóloga de larga duración que no protege contra otros serotipos y puede provocar dengue hemorrágico (Heymann, 2008), cuya frecuencia depende de la cepa del virus durante una epidemia (Cortés, Gómez y Ocazonez, 2007).

2. Distribución. Se encuentra con mayor frecuencia en áreas tropicales y subtropicales (Centers for Disease Control y Prevention, 2012).

3. Reservorio. Los virus son perpetuados en un ciclo que incluye al ser humano y al mosquito *Aedes Aegypti* en centros urbanos de clima tropical (Berger, 2015).

4. Modo de transmisión. Por la picadura de mosquitos infectantes, principalmente *Aedes Aegypti*. Esta es una especie hematófaga, con mayor actividad de picadura dos horas después de la puesta del sol y varias horas antes del amanecer (Heymann, 2008).

5. Período de incubación. De 9 a 12 días, por lo común de 5 a 7 días (Dirección General de Promoción de la Salud, 2016)

6. Período de transmisibilidad. No se transmite directamente de una persona a otra, los enfermos suelen infectar a los mosquitos desde poco antes de terminar el período febril, un promedio de tres a cinco días. El mosquito se vuelve infectante 8 a 12 días después de alimentarse con sangre virémica y permanece así el resto de su vida (Heymann, 2008).

- **Chikungunya.**

1. Descripción. Enfermedad viral febril auto limitada caracterizada por artralgia o artritis, principalmente en la muñeca, rodilla, tobillo y pequeñas articulaciones de las extremidades que duran de días a meses. Después de la picadura de un mosquito infectado, generalmente los individuos presentarán síntomas tras un período de incubación, sin embargo, no todos desarrollan síntomas. Estudios serológicos indican que entre el 3% y el 28% de las personas con anticuerpos para el virus de chikungunya tienen infecciones asintomáticas (Centers for Disease Control and Prevention, 2011).

En la Figura 2.2., se muestra la cronología de la replicación viral en relación a los signos clínicos y biológicos de la enfermedad.

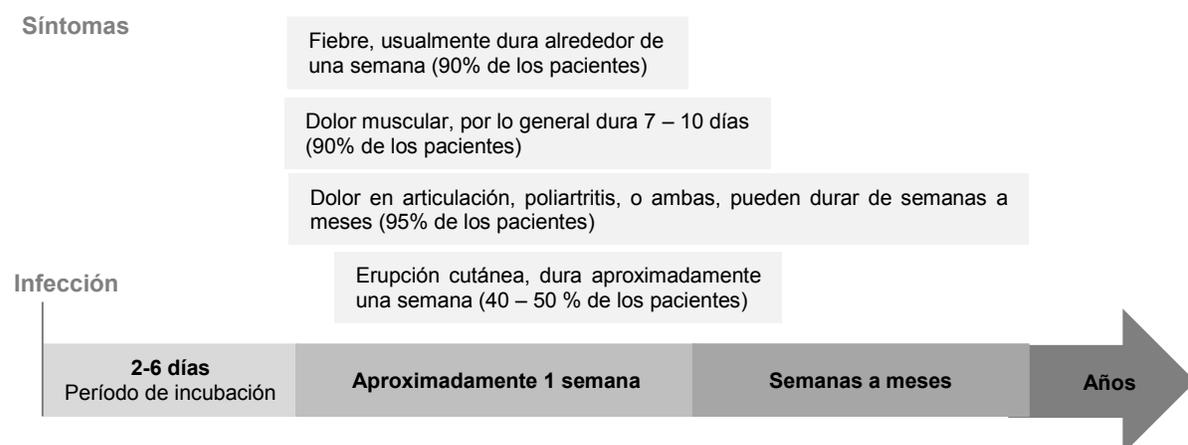


Figura 2.2. Cronología de los síntomas e infección (Weaver y Lecuit 2016).

2. Distribución. En México el primer caso importado de Fiebre Chikungunya del Caribe fue identificado en junio de 2014. En tanto que el primer caso autóctono se detectó en septiembre, sumándose un total de 222 casos al final de este año. Durante el 2015 el número de casos registrados fue 12,588 casos confirmados en 28 entidades federativas, lo que evidenció la rápida diseminación del virus (Dirección General de Epidemiología, 2016).

3. Reservorio. Durante el período epidémico, los seres humanos sirven como reservorio del virus, mientras que en los períodos inter-epidémicos un número de

vertebrados han sido implicados como reservorios en la región africana, los cuales incluyen monos, roedores y aves (Directorate General of Health Services, 2016).

4. Modo de transmisión. Existen dos vectores principales para el virus: *Aedes Aegypti* y *Aedes albopictus* (Weaver y Lecuit 2016).

5. Período de incubación. En los humanos picados por un mosquito infectado, los síntomas de enfermedad aparecen generalmente después de un período de incubación intrínseca de tres a siete días (rango: 1–12 días) (Figura 2.3.) (Centers for Disease Control and Prevention, 2011).

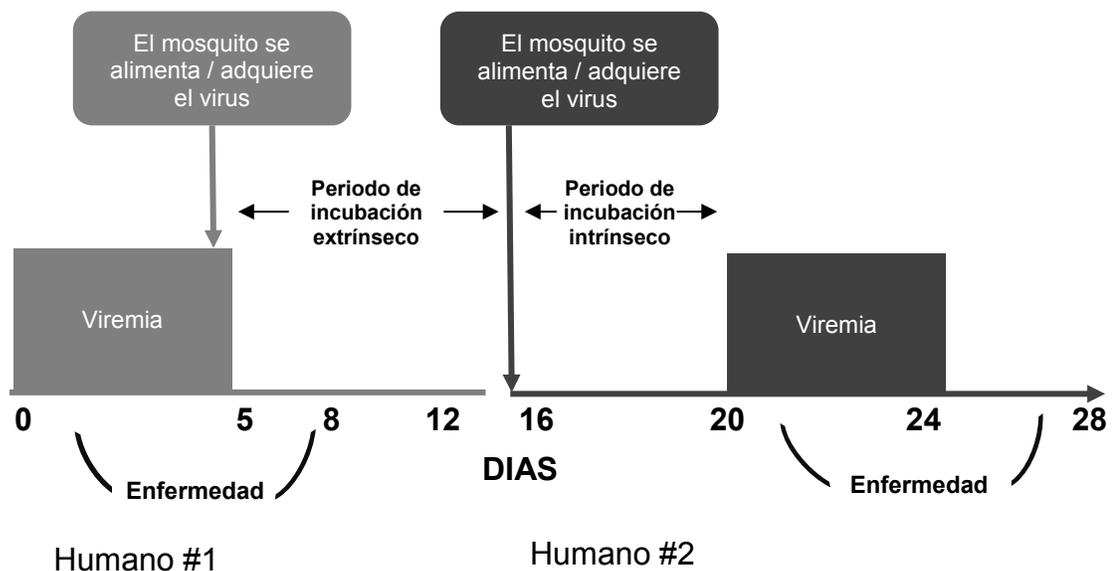


Figura 2.3. Períodos de incubación extrínseca e intrínseca (Pan American Health Organization, 2011)

6. Período de transmisibilidad. Los mosquitos adquieren el virus a partir de un huésped virémico. Después de un periodo promedio de incubación extrínseca de 10 días (Figura 2.3.), el mosquito es capaz de transmitir el virus a un huésped susceptible, como a un ser humano. Las personas con infección aguda por el virus con manifestaciones clínicas o asintomáticos, pueden contribuir a la diseminación de la enfermedad si los vectores que transmiten el virus están presentes y activos en la misma zona (Centers for Disease Control and Prevention, 2011).

- **Rickettsiosis**

Rickettsiosis es un grupo de enfermedades clínicamente similares, causadas por rickettsias semejantes. Son transmitidas por garrapatas de cuerpo duro (Acari: Ixodidae), la epidemiología de cada enfermedad rickettsial depende la distribución geográfica y las actividades estacionales de los vectores, así como de los comportamientos humanos (Centers for Disease Control y Prevention, 2016).

Los casos de brotes epidemiológicos que se presentan en el área geográfica analizada son de tipo Fiebre Maculosa de las Montañas Rocosas también llamada Fiebre Manchada, por lo que la información presentada es del tipo mencionado.

1. Descripción. Enfermedad caracterizada por fiebre moderada o alta de comienzo repentino, que, por lo general dura de dos a tres semanas en enfermos no tratados, malestar general notable, mialgias profundas, cefalalgias intensas, escalofríos e hiperemia conjuntival. Entre el tercero y el quinto días suele aparecer erupción maculopapulosa en las extremidades, que pronto abarca las palmas de las manos y las plantas de los pies, y se propaga rápidamente a casi todo el cuerpo. La tasa de letalidad es de 13 a 25% en caso de no administrarse el tratamiento específico (Heymann, 2008).

2. Distribución. Durante el 2015 se notificaron 272 casos en 24 estados siendo los más afectados Sonora, Sinaloa, Nuevo León y Morelos que concentraron el 49% de los casos del país. La tasa registrada fue de 0.22 por 100 mil habitantes. El grupo de edad más afectado fue el de 50 a 59 años con una tasa de 0.32 por 100 mil habitantes (Dirección General de Epidemiología, 2016).

3. Reservorio. Las rickettsias pueden transmitirse a los perros, a algunos roedores y a otros animales, cuyo principal vector y reservorio es la garrapata *Rhipicephalus Sanguineus* (Centers for Disease Control y Prevention, 2016).

4. Modo de transmisión. Generalmente, por la picadura de una garrapata infectada. La garrapata tiene que estar adherida durante varias horas (de cuatro a seis) y

succionar sangre para que las rickettsias se reactiven e infecten a la persona. La contaminación de abrasiones en la piel o en las mucosas por tejidos de la garrapata aplastada o sus heces, también puede causar infección (Heymann, 2008).

5. Período de incubación. De 2 a 21 días (Kansas Department of Health and Environment, 2016).

6. Período de transmisibilidad. No se transmite de persona a persona, el vector se mantiene como reservorio de por vida (Kansas Department of Health and Environment, 2016).

- **Diarrea Aguda (DA)**

1. Descripción. La diarrea es el paso de excrementos inusualmente sueltos o acuosos, por lo menos tres veces en un período de 24 horas (WHO, 2005). En 2005, la Organización Mundial de la Salud revisó trece de las guías para el tratamiento de la diarrea aguda y estableció los objetivos del tratamiento como: prevenir / tratar la deshidratación, prevenir problemas nutricionales y reducir la duración y gravedad del episodio diarreico. Estos objetivos pueden lograrse mediante el uso adecuado de terapia de rehidratación oral y terapia de rehidratación intravenosa, el mantenimiento de una ingesta de alimentos adecuada y, en algunos casos, el uso sensato de medicamentos sintomáticos (antipiréticos, analgésicos y antieméticos), zinc, antisecretores, probióticos y antibióticos (Brandt, Antunes y Pontes da Silva. 2015)

2. Distribución. La diarrea afecta a la mayoría de los seres humanos en algún momento de sus vidas. Se estima que anualmente, se producen 2 mil millones de casos de enfermedades diarreicas y 2,5 millones de muertes por enfermedades relacionadas con la diarrea alrededor del mundo (Carlson, Rose y Gelinas, 2016)

3. Modo de transmisión. Las diarreas infecciosas se transmiten por contacto directo (manos sucias) o indirecto (ingestión de agua o alimentos contaminados) (Medecins Sans Frontieres, 2016).

- **Rinofaringitis**

1) Descripción.

Este término implica el resfriado común y otros trastornos causados por la rinitis viral aguda. Es causada generalmente por virus (rinovirus, coronavirus, parainfluenza, entre otros). Tiene un periodo de incubación de dos a cinco días y puede comenzar con dolor de garganta, secreción nasal, congestión nasal, estornudos, tos seca y fiebre de diversa intensidad, generalmente más alta en niños menores de cinco años. Algunos pacientes con esta infección no tienen fiebre y ciertos tipos de virus pueden causar diarrea (Pitrez y Pitrez, 2003).

2) Distribución.

Es una enfermedad universal, entre el tercer y quinto día que es más sintomática al igual que contagiosa, siendo más frecuentes en los trópicos en épocas lluviosas. La rinofaringitis aguda es el tipo más frecuente de infección de las vías respiratorias superiores en la infancia. Los niños menores de cinco años pueden tener entre cinco y ocho episodios al año (Heymann, 2008).

3) Modo de transmisión.

Puede ser por contacto directo, el virus permanece vivo en la piel durante al menos dos horas; contacto indirecto, los virus pueden sobrevivir en superficies durante varias horas, otro modo es por inhalación de partículas virales (The Montana Department of Public Health and Human Services, 2013).

2.3. Sistemas de Información Geográfica (GIS)

Un GIS puede definirse 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 (Loslier, 2016).

En esta sección se aborda la relevancia del uso de GIS en el área de la salud, de igual manera se desarrolla el significado de la representación espacio temporal.

2.3.1. Sistemas de Información Geográfica (GIS) como herramienta dentro de la Salud Pública.

La mayoría de los datos epidemiológicos tienen una referencia de lugar y tiempo, los GIS son tecnología innovadora e ideal para manejar este tipo de información (Nobre y Carvalho, 2016). Dentro de la obra de los autores Kreis, et al. (2013) se menciona cómo esta herramienta se utiliza en la vigilancia de enfermedades para examinar enlaces espaciales y en investigaciones de brotes ayudando en la identificación de fuentes.

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 geográfica de las políticas de salud (distribución, entrega, acceso a servicios de salud y población) (Yagoub, 2011).

Dentro de los GIS se encuentran dos tipos principales de datos, están las bases de datos espaciales que contienen a los datos de localización, los cuales describen la geografía de las características de la superficie terrestre (forma, posición), y la base de datos de atributos, cuya información describen características o cualidades de los datos espaciales, como es el uso de la tierra y tipo de suelo, por mencionar algunas (Loslier, 2016).

Estos sistemas de información han reunido los métodos de investigación y técnicas analíticas de la geografía médica y la epidemiología espacial. En las últimas décadas se ha observado un aumento en el número de estudios epidemiológicos que los emplean, particularmente en áreas con variabilidad en los estados salud, disponibilidad de recursos y comportamientos relacionados con la salud (Musa *et al.*, 2013).

Entre los diversos estudios epidemiológicos que emplean GIS, se encuentra el de Alias *et al.* (2014), cuyo objetivo es determinar los cambios en la distribución espacial de la malaria en Malasia peninsular desde el año 2000-2009; por otro lado esta Gammino *et al.* (2014) quienes usan receptores de GPS portátiles para poner a prueba la aplicación del análisis geoespacial a las actividades de vacunación contra la poliomielitis en el norte de Nigeria. También se encuentra Elebead *et al.* (2012), quienes desarrollan una base de datos, con el uso de GIS, para examinar la distribución de casos de cáncer, con el objetivo de producir mapas base para la implementación de programas de control en Gezira, Sudan.

A. Representación Espacio-temporal

Un componente importante para los estudios epidemiológicos y la vigilancia de la salud pública es la interacción espacio-tiempo, ya sea entre eventos de salud o entre eventos de salud y variables ambientales (Nobre y Carvalho, 2016).

La palabra espacial se refiere a los datos asociados con la ubicación geográfica, existen diversas aplicaciones que han recopilado una gran cantidad de datos espaciales, estos datos conforman conjuntos inmensos de información, de tal manera, que es necesario llevar a cabo un análisis automático o semiautomático, lo que deriva en la necesidad de métodos nuevos y eficientes para descubrir el conocimiento en grandes bases de datos espaciales (Perumal *et al.*, 2015).

Un ejemplo se encuentra en el trabajo de Zhou *et al.*, (2015)(Figura 2.4.) quienes analizaron la distribución espacio-temporal de la incidencia de cáncer de mama en Shenzhen entre 2007 y 2012, concluyendo que esta distribución orienta a la etiología del cáncer para la prevención primaria, además de una base científica para la asignación adecuada de los recursos de salud. Otro ejemplo se puede observar en la investigación de Liu *et al.* (2015)(Figura 2.5.), en la cual, exploraron las relaciones entre el clima y una enfermedad transmisible denominada HFMD, utilizando el modelo espacio-temporal, el cual, además de reconocer la tendencia espacial y temporal de la

enfermedad, permite realizar predicciones y orientar la formulación e implementación de estrategias apropiadas para prevenir y controlar dicha enfermedad.

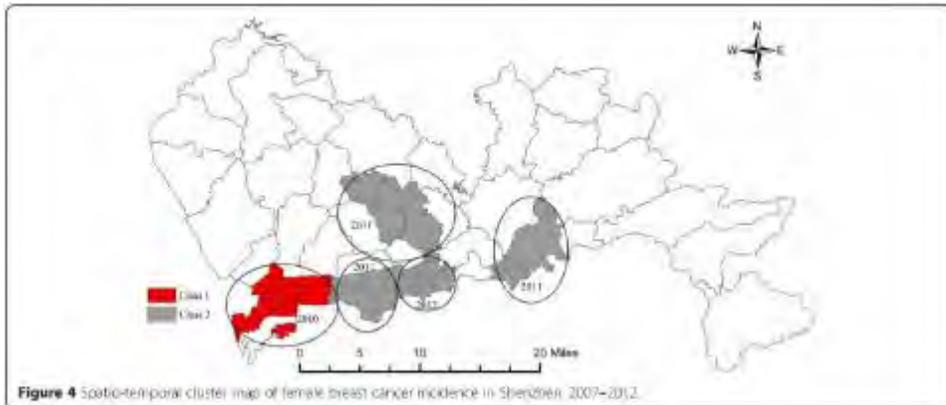


Figura 2.4. Ejemplo de una representación espacio-temporal (Zhou et al., 2015)

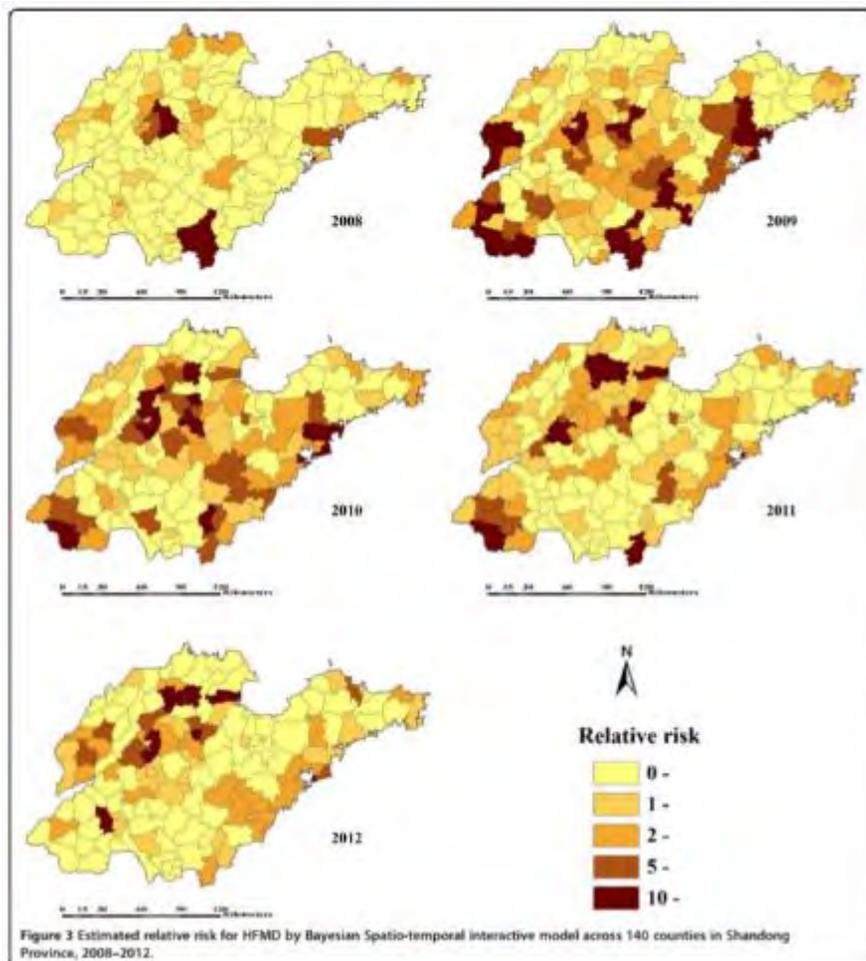


Figura 2.5. Ejemplo de representación espacio-temporal (Liu et al., 2015)

2.4. Minería de Datos

Cuando se tiene una gran cantidad de información, es necesario extraer reglas de asociación ocultas en las bases de datos para adquirir conocimiento útil y comprensible, previamente desconocido, a este proceso de extracción se le conoce como “*minería de datos*” (Huang, 2013).

Las técnicas de minería de datos se han mostrado útiles para las aplicaciones espaciotemporales y, son fundamentales para las organizaciones que toman decisiones basadas en grandes conjuntos de datos espaciales y espaciotemporales (Shekhar *et al.*, 2015). La minería de datos espaciales es una herramienta que permite la base teórica y metodológica para la identificación de patrones entre los datos, estos patrones son la herramienta principal en el análisis de los datos y la realización de proyecciones para la toma de decisiones (Cangrejo Aljure y Agudelo, 2011).

A continuación, se define la minería de datos y sus principales fundamentos. Dentro de ese apartado se presenta la importancia de la minería de datos como herramienta dentro del ámbito de la salud pública y el papel que tiene como parte del proceso de descubrimiento de conocimiento presente en las bases de datos (*Knowledge Discovery in Databases o KDD*). De igual forma, se revisan los modelos, tareas y técnicas más comúnmente aplicadas en la minería de datos.

2.4.1. Definición

En la literatura existen varias definiciones para el término minería de datos, en la siguiente matriz se realiza una comparación de algunas de ellas (tabla 2.1).

Autor	Diferencia	Semejanza
Hand, Mannila y Padhraic (2001)	Tiene el propósito de resumir los datos de forma novedosa que sea comprensible y útil para el propietario de los datos.	Análisis de una gran cantidad de datos, con el fin de encontrar relaciones insospechadas.
Rygielski, Wang y Yen (2002)	La definen como una capacidad de búsqueda de datos mediante algoritmos estadísticos	Se utiliza para descubrir patrones y correlaciones.

Witten y Frank (2005)	Resaltan la importancia de que los patrones descubiertos deben de conducir a algún beneficio, generalmente económico.	Es un proceso de descubrir patrones en los datos, y que los datos se presentan en grandes cantidades.
Sumathi y Sivanyam (2006)	Es el descubrimiento eficiente de información valiosa y oculta	La información valiosa y oculta se encuentra en una gran colección de datos.

Tabla 2.1 Matriz comparativa para las definiciones de Minería de Datos

Por la inexistencia de la definición exacta para el término, Hernández, Ramírez y Ferri (2004) resumen el objetivo de la minería de datos basándose en distintas definiciones como el proceso de convertir datos extraídos de grandes volúmenes de datos en conocimiento útil y previamente desconocido.

2.4.2. Proceso de Descubrimiento de Conocimiento en Base de Datos.

El descubrimiento de conocimiento presente en las bases de datos (*Knowledge Discovery in Databases o KDD*) es el proceso no trivial de encontrar e interpretar patrones a partir de datos, implicando la aplicación repetida de métodos específicos de minería de datos o algoritmos y la interpretación de los patrones generados por estos algoritmos (Becerra-Fernandez, Zanakis y Walczak, 2002; Shafique y Qaiser, 2014). Especificando, KDD se refiere al proceso general de descubrir conocimientos útiles a partir de los datos, y la minería de datos se refiere a un paso esencial en este proceso como se puede observar en la Figura 2.6. (Fayyad, Piatetsky-shapiro y Smyth, 1996).

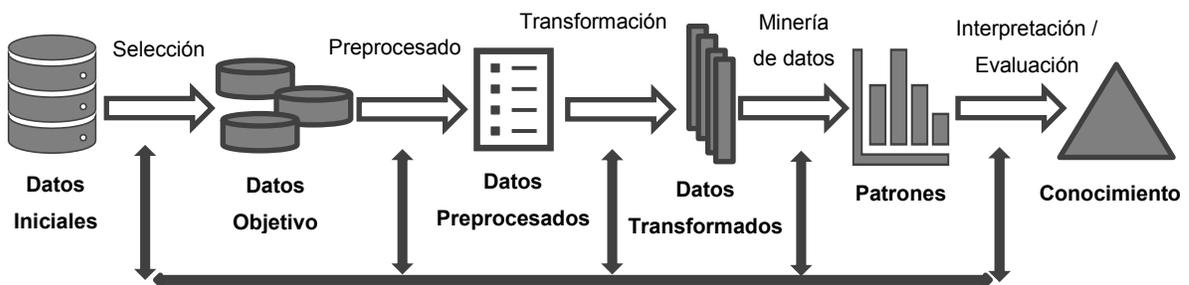


Figura 2.6. Etapas para el proceso de extracción del conocimiento (Fayyad, Piatetsky-Shapiro y Smyth, 1996)

2.4.3. Modelos y tareas de minería de datos.

La minería de datos está constituida por modelos *predictivos* y *descriptivos*, en la práctica, la mayoría de las aplicaciones de KDD requieren cierto grado de modelado tanto predictivo como descriptivo. La mayoría de los métodos se basan en técnicas comprobadas de aprendizaje automático, reconocimiento de patrones y estadísticas, como clasificación, regresión, agrupación (Clustering), etc.; la representación de un modelo que se utiliza por un método particular viene normalmente compuesto por opciones como splines, polinomios, funciones de núcleo e imagen, función booleana, por mencionar algunos (Fayyad, et al., 1996), en la figura 2.7 se desarrollan los modelos de conocimiento (Gibert, Sánchez-Marrè y Codina, 2006).

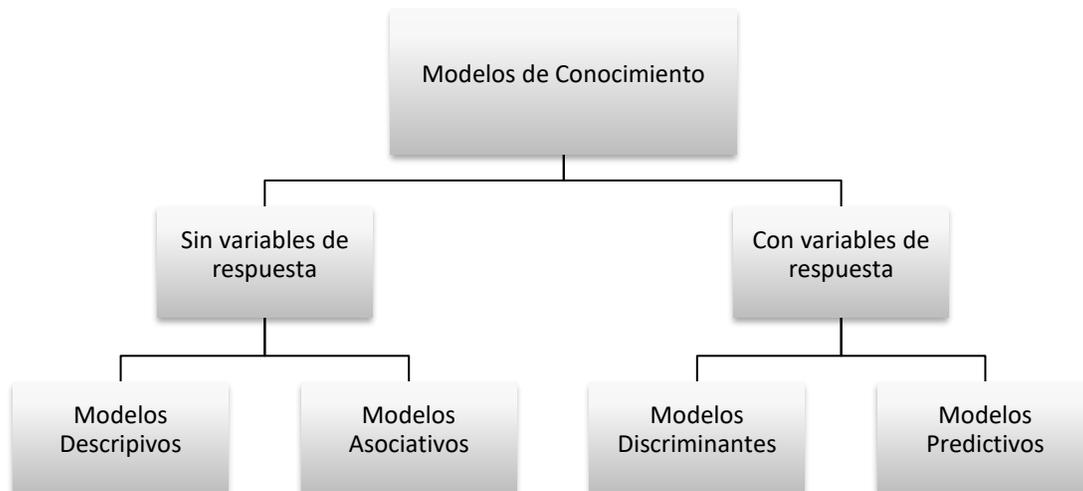


Figura 2.7. Modelos de Conocimiento (Gibert, Sánchez-Marrè y Codina, 2006).

- Modelo Predictivo

Se emplea cuyo el objetivo es estimar valores futuros de variables de interés basándose en los valores previos disponibles mediante el proceso de tomar patrones descubiertos en la base de datos (Han, Kamber y Pei, 2012).

- Modelo Descriptivo

Descubrir patrones o relaciones describiendo los datos, se centra en la comprensión del proceso subyacente de la generación de datos, se interpreta como un reflejo de la realidad (Fayyad, et al., 1996; Dueñas, 2009).

- Aprendizaje Supervisado

Plantea modelar un sistema en base a datos de entrada y de salida del mismo, se debe disponer de suficiente cantidad de información para que el modelo a construir permita aprender sobre la información provista, y así representar el comportamiento deseado, algunos ejemplos de tareas de este tipo de aprendizaje son regresión y clasificación, con algoritmos que incluyen regresión logística y máquinas de vectores de soporte (Shickel *et al.*, 2017).

- Aprendizaje no Supervisado

Tiene el objetivo es capturar características relevantes de los datos presentados, en este tipo de aprendizaje no existe una distinción entre datos de entrada y de salida de un sistema, ya que en este caso no se busca aprender su comportamiento sino aprender propiedades interesantes sobre la distribución de x en sí, algunos ejemplos de tareas de aprendizaje no supervisado es el agrupamiento y la estimación de la densidad (Shickel *et al.*, 2017).

- Tareas de minería de datos

En el proceso de la minería de datos, se encuentran distintas tareas específicas que corresponden a un tipo de problema determinado (Rygielski, Wang y Yen, 2002).

La Figura 2.7. muestra una representación general de los modelos descritos anteriormente y las diversas tareas que comprenden, las cuales se describen a continuación de manera breve:

- Regresión.

Tanto Fayyad, Piatetsky-Shapiro y Smyth (1996a;1996b) y Riquelme, Ruiz y Gilnert (2006) coinciden que el propósito de la regresión es relacionar un dato con un valor real de una variable de predicción, si se dispone de una variable definida se trata de regresión simple, si se cuentan con más variables es regresión múltiple (figura 2.8).

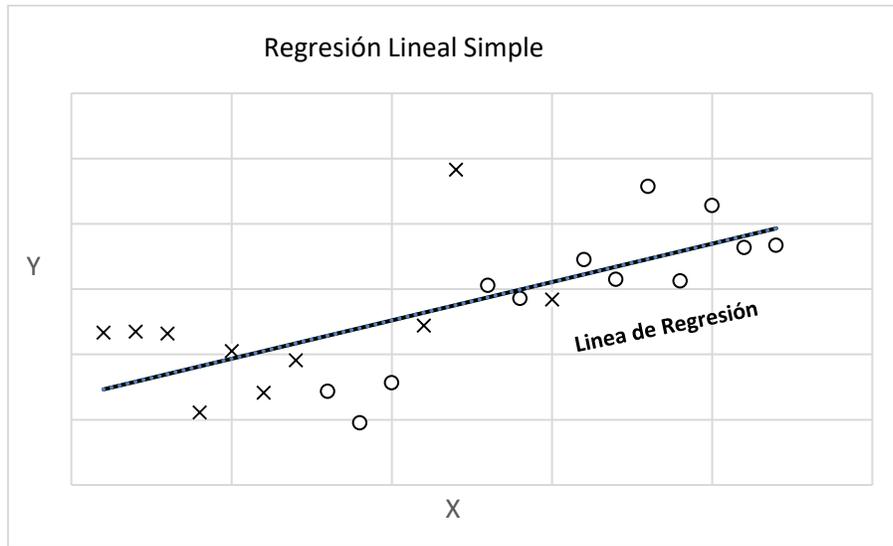


Figura 2.8. Representación de la Regresión lineal simple (Fayyad, Piatetsky-shapiro y Smyth, 1996)

- Clasificación.

Se refiere a asignar o clasificar un elemento de datos en una de varias clases categóricas predefinidas (Fayyad, Piatetsky-shapiro y Smyth, 1996; Miller y Han, 2009) (figura 2.9).

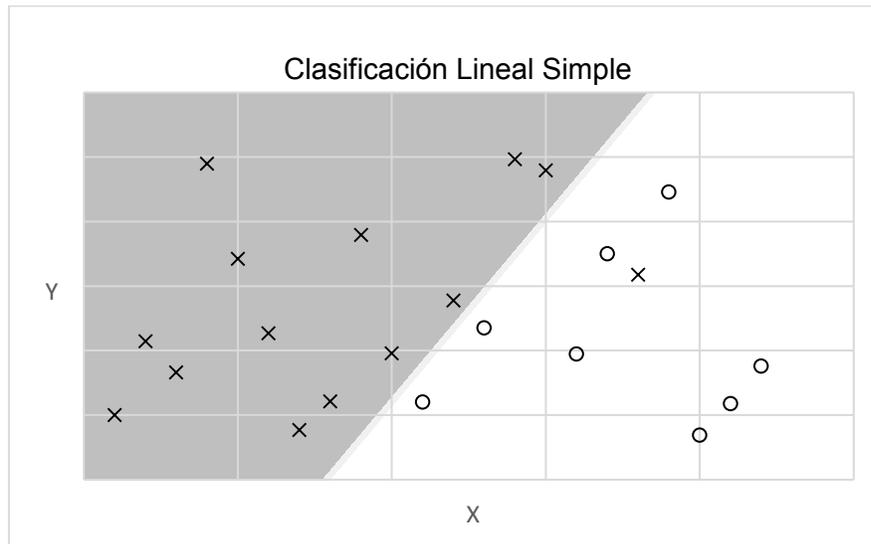


Figura 2.9. Representación de la Clasificación lineal simple (Fayyad, Piatetsky-shapiro y Smyth, 1996)

- Correlaciones.

Se utiliza cuando se requiere determinar el grado de similitud de los valores de dos variables numéricas, se mide mediante el coeficiente de correlación R , el cual es un valor real comprendido entre -1 y 1 (Hernández, Ramírez y Ferri, 2004).

- Reglas de Asociación.

Las reglas de asociación son similares a la clasificación, con la excepción de que pueden predecir cualquier atributo además de la clase, lo que les permite predecir combinaciones de atributos. Es un tipo de relación de dependencia expresada $X \Rightarrow Y$ (c%, r%) donde X y Y son conjuntos disjuntos de elementos de una base de datos. Las diversas reglas de asociación expresan distintas regularidades subyacentes al conjunto de datos, y generalmente predicen cosas diferentes, de un pequeño conjunto de datos pueden derivarse varias reglas de asociación, por lo que el interés se limita a gran volúmenes de datos con alta precisión, siendo la precisión, también llamada confianza (c%), la probabilidad condicional $P(Y|X)$, en otras palabras la proporción de transacciones que contienen X que también contienen Y; ó el número de datos que predice correctamente; y la cobertura, también llamado soporte (r%), la proporción de transacciones que contienen la unión de X y Y, $P(X \cup Y)$; ó el número de datos hallados correctamente (Witten y Frank, 2005; Miller y Han, 2009).

- Agrupamiento (Clustering).

Es una tarea descriptiva de la minería de datos con el objetivo de establecer o determinar un conjunto finito de categorías o clúster que describa los datos, dichas categorías se definen mediante agrupaciones de datos que cumplen métricas de similitud o modelos de densidad de probabilidad (Fayyad, Piatetsky-shapiro y Smyth, 1996a, 1996b; Miller y Han, 2009) (figura 2.10).

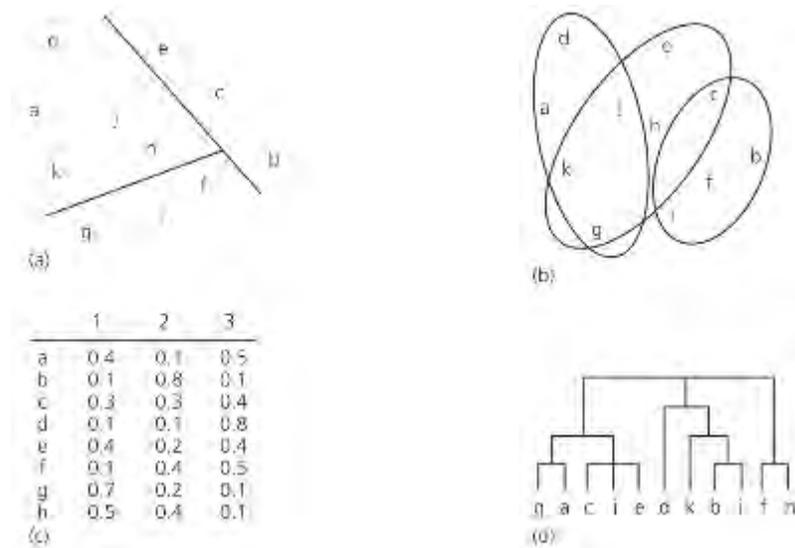


Figura 2.10. Distintas formas de representar clúster (Witten y Frank, 2005)

2.4.4. Técnicas de minería de datos

Las técnicas de minería de datos se utilizan con el propósito de encontrar y descubrir patrones estructurales en los datos, como herramienta para ayudar a explicar esos datos y realizar predicciones a partir de ellos (Witten y Frank 2005), 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 (Buczak, et al., 2014).

Existe una amplia variedad de técnicas de minería de datos. En la tabla 2.2, se muestra la relación que existe entre las tareas y las técnicas de minería de datos más utilizadas en investigaciones dentro del área de la salud (Crockett, Johnson y Eliason, 2017), la elección de las técnicas se determina considerando el tipo de datos, objetivo deseado, comprensibilidad y el tratamiento a la incertidumbre que genere el propio modelo (Riquelme, Ruiz y Gilnert, 2006).

Nombre	Predictivo		Descriptivo		
	Clasificación	Regresión	Agrupamiento	Reglas de Asociación	Correlaciones
Árboles de decisión ID3, C4.5, C5.0	✓				
Arboles de decisión CART	✓	✓			
Regresión lineal y logarítmica		✓			✓
Regresión logística	✓			✓	
K-medias			✓		
Naïve Bayes	✓				
Vecinos más próximos	✓	✓	✓		
Redes Neuronales	✓	✓	✓		
A priori				✓	

Tabla 2.2. Relación existente entre las tareas y técnicas de minería de datos (Adaptado de Hernández, Ramírez y Ferri, 2004)

2.4.5. Minería de Datos como herramienta dentro de la Salud Pública.

La minería de datos es una de las áreas de investigación en constante crecimiento, las distintas tareas y técnicas existentes, para predecir el comportamiento y las tendencias en los datos, permiten tomar decisiones más precisas basadas en el conocimiento, dichas técnicas se han utilizado con éxito en diversas áreas, incluyendo la salud pública (Shafique *et al.*, 2016).

De acuerdo con Crockett, Johnson y Eliason (2017), existe una gran variedad de aplicaciones para la que se puede utilizar, en su investigación, desarrollan minería de datos para evitar readmisiones hospitalarias, predecir el riesgo de la población de pacientes, mejorar la información de atención primaria, impulsar una mejor administración de las poblaciones en las clínicas del sistemas de salud, entre otras.

Por otro lado, Akay, Dragomir y Erlandsson (2015), presentan un marco de análisis para la extracción inteligente de conocimiento de las redes sociales, el cual se centra en el sentimiento positivo y negativo, al igual que en los efectos secundarios del tratamiento, con el propósito de brindar información de manera rápida y actualizada a hospitales, personal médico e industria farmacéutica, sobre la efectividad (o ineficiencia) de tratamientos; afirman que la extracción de conocimiento en las redes

sociales permite mejorar los resultados de atención médica, al igual que una reducción de costos.

Shafique *et al.* (2016), consideran necesario aplicar minería de datos en el sector salud, afirman que es una tarea importante debido a que su objetivo es resolver problemas de salud del mundo real.

2.5. Análisis de Datos

En esta sección se presentan los temas principales para el análisis de datos, en específico la estadística descriptiva.

2.5.1. Estadística Descriptiva

Es la primera tarea a realizar en el análisis de los datos de cualquier investigación, debido a que permite conocer, representar y cuantificar el comportamiento de un conjunto de datos (Montgomery y Runger, 2012). Fernando Quevedo (2011) realiza un estudio de las medidas estadísticas en datos agrupados y no agrupados, en la investigación de la salud, y presenta desarrolla explícitamente las siguientes herramientas:

- Medidas estadísticas en datos no agrupados

$$\text{Media: } \bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

$$\text{Desviación: } Y_i - \bar{Y}$$

Mediana: Sobre un conjunto de n datos ordenados $x_{(1)}, x_{(2)}, x_{(3)}, \dots, x_{(n)}$ la mediana es

$$x_{\left(\frac{n+1}{2}\right)} \text{ si } n \text{ es impar} \qquad \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} \text{ si } n \text{ es par}$$

$$\text{Varianza: } \sigma^2 = \frac{\sum_i (y_i - \mu)^2}{N} \qquad s^2 = \frac{\sum_i (y_i - \bar{y})^2}{y_i - 1}$$

$$\text{Desviación estándar: } s = \sqrt{\frac{\sum_i (y_i - \bar{y})^2}{y_i - 1}} \qquad \text{Coeficiente de variación: } \frac{s}{\bar{y}} \times 100$$

- Medidas de tendencia central y de dispersión en datos agrupados

$$\text{Media: } \bar{y} = \frac{\sum_{i=1}^n y_i n_i}{n}$$

$$\text{Mediana: } LI + \frac{\left[\frac{n}{2} - N_{i-1}\right]}{n_i} A_i$$

$$\text{Varianza: } s^2 = \frac{\sum_i (y_i - \bar{y})^2 n_i}{n-1}$$

$$\text{Desviación estándar: } s = \sqrt{\frac{\sum_i (y_i - \bar{y})^2 n_i}{n-1}}$$

$$\text{Percentil } j \text{ (P}_j\text{)} : LI + \frac{\left[\frac{jn}{100} - N_{i-1}\right]}{n_i} A_i$$

2.6. Software

Al trabajar con grandes conjuntos de datos, es necesario llevar a cabo la implementación de software que permita automatizar las fases de análisis, realizarlos con mayor rapidez y que inhabilite el error humano.

En la tabla 2.3 se exponen algunos de los programas que se pueden utilizar en investigaciones donde es necesario aplicar GIS y Minería de datos.

Software	Descripción Breve	Características	Ventajas	Desventajas	Aplicaciones/ Usos
ArcGIS	Es un completo sistema que permite recopilar, organizar, administrar, analizar, compartir y distribuir información geográfica.	<ul style="list-style-type: none"> • Potente GIS de escritorio • Crear y Diseñar mapas y escenas 3D • Análisis • Administrar datos geográficos • Imágenes • Ampliar el alcance • GIS 3D 	<ul style="list-style-type: none"> • Escalable • Aplicaciones para todo el mundo • Mapas listos para usar • Autoservicio • Administración • Seguro y de confianza 	<ul style="list-style-type: none"> • ArcGIS obliga a conectarse a una licencia para realizar determinadas operaciones de geoprocésamiento. • La instalación es un proceso más extenso en tiempo, unido a que además se necesita instalar diferentes extensiones y realizar la autorización del software desde <i>ArcGIS Administrator</i>. 	<ul style="list-style-type: none"> • Recopilar, organizar, administrar, analizar, compartir y distribuir información geográfica. • Crear, compartir y utilizar mapas inteligentes <ul style="list-style-type: none"> • Crear y administrar bases de datos geográficas • Resolver problemas con el análisis espacial

<p>QGIS</p>	<p>GIS</p>	<ul style="list-style-type: none"> • Se puede ver y sobreponer datos vectoriales y ráster en diferentes formatos y proyecciones sin convertir a un formato interno o común. 	<ul style="list-style-type: none"> • Software OpenSource • Versatilidad en el consumo de datos <ul style="list-style-type: none"> • Conexión servicios OGC • Procesado de imágenes de satélite (Teledetección) • Estética de etiquetado <ul style="list-style-type: none"> • Geo-codificación 	<ul style="list-style-type: none"> • Limitación en el número de archivos abiertos • Tiene varias tareas que se realizan de forma más sencilla en otros GIS 	<ul style="list-style-type: none"> • Recopilar, organizar, administrar, analizar, compartir y distribuir información geográfica. • Crear, compartir y utilizar mapas inteligentes
<p>Google Maps</p>	<p>Plataforma de mapas web desarrollada por Google, funciona mediante la superposición de fotografías vista satélite, así como desde panorámicas a nivel de calle, es un servicio que ofrece tecnología cartográfica potente y muy fácil de usar, así como información sobre empresas locales</p>	<ul style="list-style-type: none"> • Vista de mapas • Obtener información sobre un sitio <ul style="list-style-type: none"> • Obtener indicaciones • Definir direcciones a múltiples destinos • Destinos de atractivo turístico • Simplicidad • Ofrece la capacidad de hacer acercamientos o alejamientos para mostrar el mapa. • Vista previa de vuelos, hoteles y reserva en restaurantes 	<ul style="list-style-type: none"> • Fácil de utilizar. <ul style="list-style-type: none"> • Medir distancia entre dos o más puntos del mapa • Obtener o buscar coordenadas • Ofrece mucha información adicional sobre gran parte de los lugares del mundo. • Muestra rutas para llegar a diferentes desde nuestra situación. <ul style="list-style-type: none"> • Se puede imprimir, enviar o enlazar los mapas. • Buena calidad de imágenes 	<ul style="list-style-type: none"> • Requiere obligatoriamente una conexión a Internet. • Google Street View aún no está completo, ya que solo nos ofrece algunas ciudades. • Necesita de algún navegador • Depende de la conexión la velocidad • Se necesita cargarla completamente • En ciudades no muy importantes el nivel de detalle es muy bajo • Hay lugares que no aparecen • Algunas imágenes no son actuales 	<ul style="list-style-type: none"> • Análisis y consulta de información geográfica • Ejerce como GPS • Crear y compartir mapas • Dar a conocer y compartir información mediante la geografía y la visualización

<p>Google Earth</p>	<p>Es una herramienta que permite desplazarte a cualquier lugar de la Tierra para ver imágenes de satélite, mapas, relieve y edificios en 3D e incluso explorar galaxias del firmamento. Puedes recorrer contenido geográfico muy variado, guardar los lugares que visites y compartirlos con otros.</p>	<ul style="list-style-type: none"> • Interfaz en inglés, español, francés y alemán • Imágenes históricas • Datos del lecho y la superficie oceánica proporcionados por expertos <ul style="list-style-type: none"> • Viajes simplificados con posibilidad de añadir audio y grabaciones de voz • Trabaja con Sketch Up • Panel de mandos más discretos y gana espacio para la visualización de imágenes. 	<ul style="list-style-type: none"> • Fácil uso y mucha información. • Descubre lugares o revive aquellos en los que ya han estado. <ul style="list-style-type: none"> • Entrega fotografías y videos. • Versión digital del planeta. • Imágenes reales. <ul style="list-style-type: none"> • Las aplicaciones han facilitado la vida de muchos y sobre todos de los negocios. • Los medios de comunicación pueden contar la historia en tiempo real con menos problemas. 	<ul style="list-style-type: none"> • Mala calidad de las fotografías (cercanas, borrosas) • Entrega información de algunas zonas sin actualización • Muestra todo lo que hacen los usuarios comunes (incluyendo las coordenadas) • Información Fragmentaria 	<ul style="list-style-type: none"> • Análisis cartográfico • Análisis y consulta de información geográfica • Dar a conocer y compartir información mediante la geografía y la visualización
<p>MATLAB</p>	<p>Es un entorno de cálculo técnico de altas prestaciones para cálculo numérico y visualización</p>	<ul style="list-style-type: none"> • Lenguaje de alto nivel para cálculos científicos y de ingeniería • Entorno de escritorio afinado para la exploración iterativa, el diseño y la solución de problemas 	<ul style="list-style-type: none"> • Amplio soporte Matemático además de posibilidad de uso de precisión extendida en los cálculos. • Comercial, es fácil de adquirir. 	<ul style="list-style-type: none"> • Problemas eventuales de velocidad • Proceso laborioso susceptible a cometer errores en la programación de algoritmos; es conveniente depurar correctamente los códigos realizados 	<p>Para resolver clases particulares de problemas como:</p> <ul style="list-style-type: none"> • Procesamiento de señales • Diseño de sistemas de control • Simulación de sistemas dinámicos • Identificación de sistemas <ul style="list-style-type: none"> • Redes neuronales y otros.

Tabla 2.3. Matriz de Software

2.7. Análisis de Estudios Previos

Un ejemplo práctico de la implementación de minería de datos y GIS para la predicción de epidemias se encuentra en la investigación de Buczak, et al. (2015), en la cual se desarrollaron modelos para predecir alta, media y baja incidencia de Malaria en Corea del Sur, mediante la aplicación de técnicas de minería de datos. Los autores extraen las relaciones entre datos epidemiológicos, meteorológicos y socioeconómicos, estas relaciones son en forma de reglas, de las cuales el mejor conjunto de reglas es elegido automáticamente y forma un clasificador para convertirlo en un modelo de predicción de malaria.

Actualmente se pueden encontrar varios estudios relacionados, como el de Sirisena, et al. (2017) donde aplican GIS para cartografiar y evaluar la distribución espacial y temporal del dengue en Sri Lanka, con el propósito de explicar la asociación de factores climáticos con la incidencia del dengue. Recopilaron datos epidemiológicos, poblacionales y meteorológicos de la Unidad de Epidemiología, para posteriormente analizarlos mediante el software "SPSS", obteniendo como conclusión que la implementación de GIS utilizando factores de riesgo climáticos con incidencia de dengue, ayudaría a las autoridades sanitarias y al gobierno a examinar las áreas de alto riesgo y, a hacer la planificación y aplicación apropiadas de las medidas de control.

Jensen, Jensen y Brunak (2012) realizan un estudio en el cual se enfocan en cómo el descubrimiento de conocimiento de datos de un repositorio puede ayudar a comprender ciertas brechas de conocimiento y ayudar a la toma de decisiones clínicas informadas, con el propósito de promover la investigación médica y el cuidado clínico al utilizar registros electrónicos de salud, para lograr esto, presentan cuatro maneras de analizar los datos, la primera es calculando el número esperado de pacientes asociados a ciertas enfermedades, la segunda es un clasificador de árbol de decisión utilizando características clínicas, la tercera forma que presentan es agrupando pacientes según su perfil clínico y por último reuniendo un repositorio de casos con controles de coincidencia.

3. METODOLOGÍA

GIS es la tecnología ideal para manejar datos epidemiológicos (Nobre y Carvalho, 2016). La cantidad de incidencias epidemiológicas en la población se analiza comúnmente en términos de distribución geográfica, factores de riesgo clínico, características demográficas o fuentes de exposición como las redes sociales (Carroll *et al.*, 2014). En las últimas dos décadas, la capacidad para recopilar y almacenar datos ha superado la capacidad de procesar, analizar y explotar dichos datos (Witten y Frank, 2005). Un ejemplo de eso, es la cobertura y el volumen de los conjuntos de datos geográficos digitales, los cuales cada vez son más extensos al crecer constantemente (Cangrejo Aljure y Agudelo, 2011). El valor de estos conjuntos de datos, depende de la capacidad de extraer información útil, lo que provoca, que la necesidad de técnicas, eficientes y efectivas, para la extracción y análisis de los datos espacio-temporales, sea una prioridad de investigación (Compieta *et al.*, 2007).

Por lo anterior, se debe desarrollar una metodología que permita el análisis de datos epidemiológicos mediante la aplicación de GIS, considerando indispensable el apoyo e inclusión de técnicas de minería de datos.

Buczak *et al.* (2012; 2014; 2015) han presentado en sus investigaciones la aplicación adecuada de un método, objetivo y reproducible, diseñado para predecir si una alta incidencias de enfermedades (como la debida a un brote epidemiológico) ocurrirá varias semanas en el futuro, utiliza técnicas de minería de datos, y así, filtrar grandes cantidades de datos; el método que ha utilizado de manera recurrente en sus investigaciones, selecciona automáticamente aquellas asociaciones que cumplen con ciertos criterios preestablecidos, y utiliza estas selecciones como reglas para un modelo de predicción, de esta forma, limita el número de reglas y variables a aquellas que son más importantes para el pronóstico.

La precisión del modelo, depende de dos factores: El primero es que mientras existan asociaciones sólidas entre los datos del predictor y la información que se vaya a predecir, el método es capaz de construir automáticamente un modelo de predicción

con una precisión razonable; el segundo, es la calidad de los datos, entre mayor sea la calidad, más preciso será el modelo. Por lo que, si se disponen de datos de calidad razonable, se espera que el uso de este método para predecir la incidencia de altas/bajas para otras enfermedades proporcione un rendimiento similar; los modelos desarrollados por los autores son robustos y capaces de predecir con buena precisión.

Un punto que se establece en las investigaciones de Buczak *et al.* (2012; 2014; 2015), es que un método predictivo, verdaderamente riguroso, constituido por dos características que no pueden ser transgredidas; la primera, es que el método no puede desarrollarse y probarse con exactamente los mismos datos, la validación rigurosa requiere que los datos utilizados para las pruebas no sean los mismos datos utilizados en su desarrollo.

La segunda característica es que todas las variables de predicción deben recopilarse para el período de tiempo anterior y usarse para la predicción de brotes durante un período de tiempo posterior. El propósito de esta segunda característica es asegurar una predicción realista, ya que los valores de todas las variables predictoras pueden obtenerse antes de realizar la predicción para el próximo período de tiempo.

Otro factor relevante en estas investigaciones, es como los autores dividen los datos en tres subconjuntos: entrenamiento, validación y prueba; el subconjunto de entrenamiento es el que utilizan para desarrollar el modelo. Los modelos tienen ciertos parámetros que pueden ser ajustados por los desarrolladores del modelo: el mejor modelo es el que tiene el menor error en el subconjunto de datos de validación. Una vez que se establece el mejor modelo, se convierte en la versión final, y es el que se somete a prueba para evaluar su rendimiento en un conjunto de datos no vistos previamente siendo el subconjunto de pruebas.

Los autores aplican el mismo método, para la creación de modelos de predicción, a distintas enfermedades. El cambiar de enfermedad, implica cambios en el algoritmo de creación de clasificadores, siendo esto, lo que crea un nuevo modelo. Si bien la técnica de creación de modelos es similar (excepto para estos cambios algorítmicos)

para todas las enfermedades, el modelo final no es el mismo. El conjunto de variables identificadas para el modelo de cierta enfermedad, se basa en la literatura publicada sobre esa enfermedad en específico.

En este proyecto de investigación, se propone una metodología (figura 3.1) basada en el método descrito en las investigaciones Buczak *et al.* (2012; 2014; 2015), por generar modelos efectivos para predecir brotes de enfermedades.

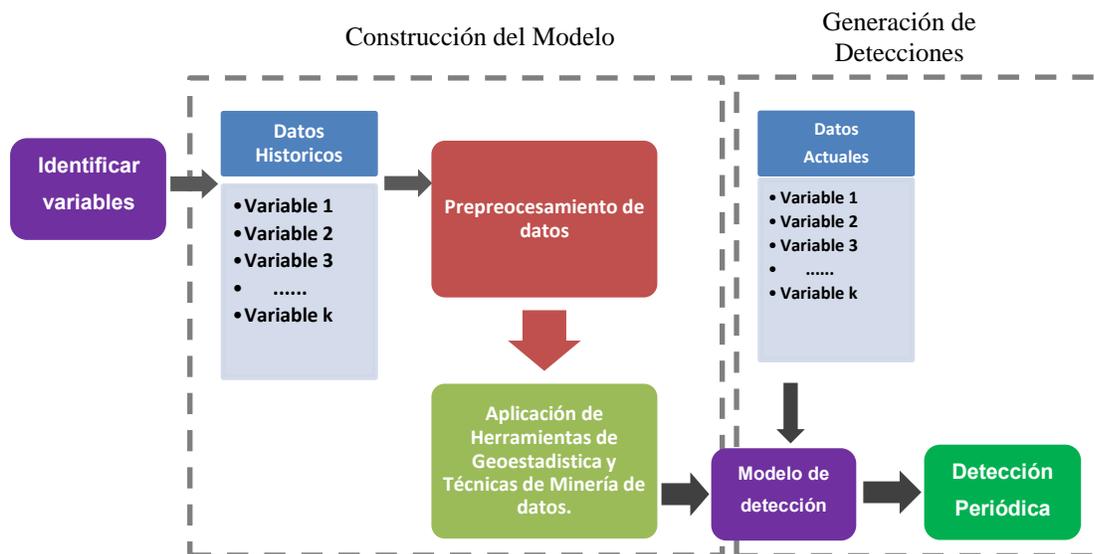


Figura 3.1. Etapas de la metodología propuesta, adaptado de Buczak et al. (2012; 2014; 2015)

A continuación, se hace una descripción de cada uno de los pasos contenidos en la metodología propuesta.

3.1. Fase I. Identificar variables relevantes para su posterior análisis.

Se realiza una revisión de la literatura reciente sobre epidemiología y acciones de prevención de la salud, con el propósito de identificar las variables que están correlacionadas con la incidencia de cierta epidemia, posteriormente se buscan las bases de datos que contengan las variables identificadas, dichas bases de datos deben de provenir de fuentes oficiales que les proporcionen validez a los datos.

Las fuentes de los datos, de donde se obtengan las variables de estudio, pueden provenir del cruce de diversas fuentes, en caso de que diversos organismos oficiales contengan bases de datos de una misma variable, se deberá realizar un análisis para elegir la base de datos que más se adecue a la investigación.

3.2. Fase II. Construcción del modelo.

Parte principal donde se encuentran los elementos de georreferenciación y minería de datos. Los datos son pre-procesados mediante estadística descriptiva 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.

Paso 1. Preprocesamiento de datos: Los datos se procesan previamente para convertirlos en la resolución espaciotemporal deseada. En este paso se estudia la calidad de los datos y establecen las operaciones de minería de datos que se pueden realizar, generalmente es la etapa más laboriosa ya que representa aproximadamente el 60% del esfuerzo total, mientras que la minería de datos solo representa el 10%.

Es un paso es fundamental, ya que establece el tipo de técnicas de minería de datos que se va a utilizar para detectar los brotes epidemiológicos, el cual, se define según los datos que se tengan.

Paso 2. Aplicación de herramientas de geoestadística y técnicas de minería de datos: Con la información lista para ser analizada se aplican las herramientas de geoestadística y técnicas de minería de datos, los resultados que se obtienen, se envían mediante informes estadísticos a la institución de salud para su validación (figura 3.2).

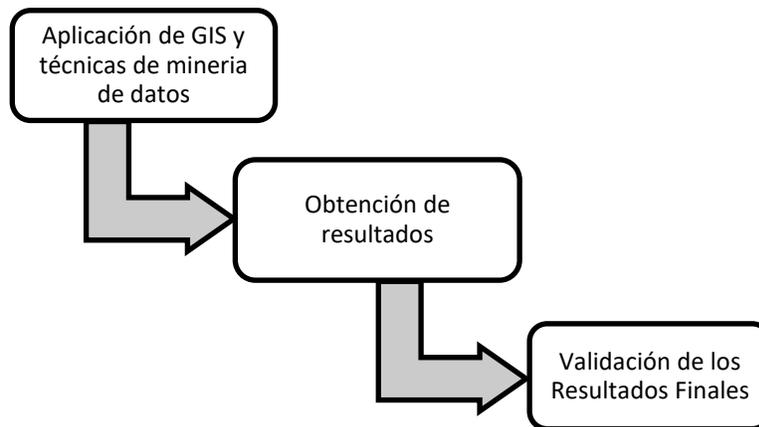


Figura 3.2. Desarrollo del paso dos de la metodología

3.3. Fase III. Generación de detecciones.

La aplicación de técnicas de minería de datos, del paso anterior, se convierte en los modelos que generan detecciones de brotes epidemiológicos (figura 3.3).

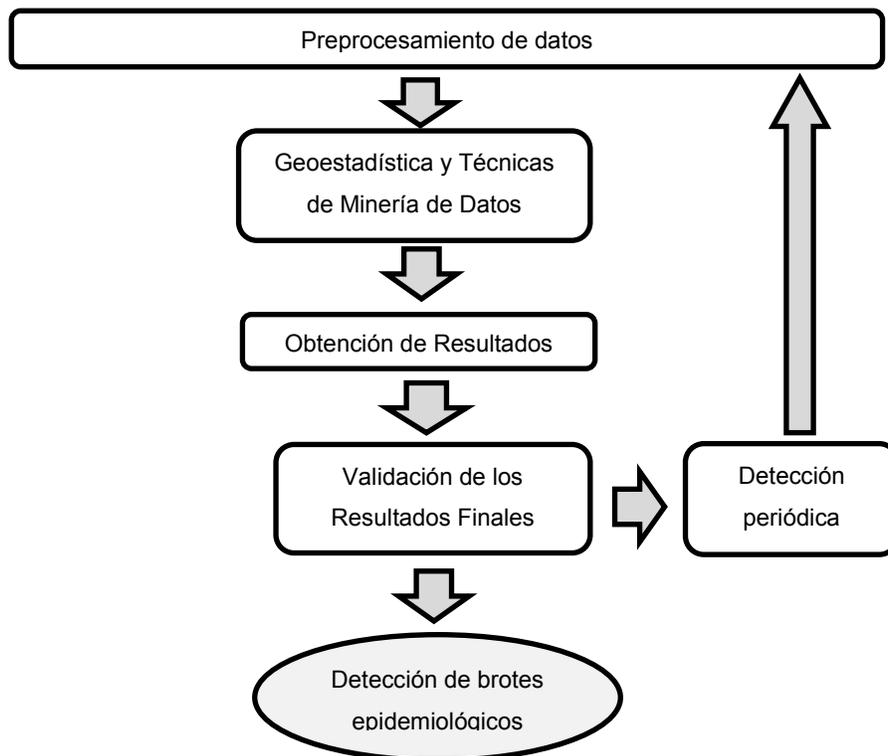


Figura 3.3. Desarrollo de pasos en la generación de detecciones

3.3.1 Detección Periódica: 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 (figura 3.4).

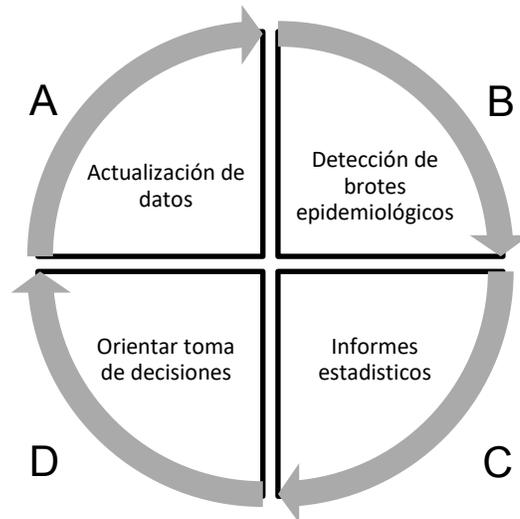


Figura 3.4. Detección periódica

A. Actualización de datos

La base de datos de la institución pública actualiza los registros epidemiológicos cada semana, por lo tanto, se tiene que estar constantemente descargando los nuevos datos registrados, con el propósito de analizar el comportamiento que están adquiriendo las enfermedades.

B. Detección de brotes

A los datos actualizados posteriormente adquiridos, se le debe realizar la metodología establecida, con el fin de emitir alertas y preparar una pronta y eficaz respuesta.

C. Informes estadísticos

Los análisis de los datos, se presenta a la institución de salud pública utilizando informes estadísticos, los cuales, independientemente de la epidemia analizada, estarán constituidos por cinco partes, que se describen a continuación:

1. Portada o Página de Presentación: Sección constituida por el título, claro y concreto, la fecha en que se entrega el reporte y, por último, nombre del autor incluyendo la afiliación institucional.
2. Índice: Establece todas las partes que constituyen del informe, al igual que el total de páginas.
3. Introducción: Es un breve resumen con la información sobre la epidemia analizada, definición de las variables, y cualquier otra información relevante a considerar para la comprensión del análisis.
4. Resultados: En esta parte se describe **qué** se ha encontrado.
 - Se presentan en forma de gráficos, se utilizarán tablas únicamente cuando sean indispensables.
 - Es la etapa donde se realiza el análisis estadístico, sea descriptivo o inferencial.
 - La información que se presente debe incluirse solo una vez, ya sea en tablas, gráficos o texto.
 - Los gráficos y/o tablas deben ser autosuficientes, no obligar al lector a recurrir al texto para comprenderlos.
 - Presentar los cálculos con el rigor necesario para validar la prueba estadística, pero no incluir información innecesaria.
 - Se presenta la interpretación de los hallazgos del estudio y realiza las recomendaciones que surjan del mismo.

5. Conclusiones

Parte final del informe, la cual contiene los resultados más importantes desarrollados de manera directa y concisa.

D. Orientar toma de decisiones

El propósito de la entrega de informes estadísticos, es orientar y fundamentar las acciones de salud, a través de la planeación y evaluación de los diferentes problemas mediante la investigación.

4. IMPLEMENTACIÓN

Esta sección se divide en tres partes: la primera, explica las fuentes oficiales de donde se obtuvieron las variables relevantes utilizadas en la investigación. En la segunda parte, se desarrolla la estructura de los datos para ser preprocesados, se detalla cómo se obtuvieron los resultados de la aplicación de las técnicas de minería de datos y las herramientas del software ArcGIS, dichos resultados serán expuestos en el capítulo 5.

4.1. Fase I. Identificar variables relevantes para su posterior análisis.

Las variables establecidas para la entrada de la construcción del modelo incluyen datos epidemiológicos y ambientales (tabla 4.1), ya que se han mostrado útiles en investigaciones previas.

Una variable importante para la carga de casos de cualquier epidemia en un año dado es la cantidad de medidas de control proporcionadas.

Datos	Fuente
Datos Climatológicos	Base de Datos Climatológica Nacional (CLICOM) http://clicom-mex.cicese.mx/mapa.html
	Estaciones Meteorológicas Automáticas (EMAS) smn.cna.gob.mx/es/emas
Datos Epidemiológicos	Base de datos de la Institución de Salud Pública
	Secretaría de Salud https://www.gob.mx/salud/acciones-y-programas/direccion-general-de-epidemiologia-boletin-epidemiologico
Datos Geográficos	Instituto Nacional de Estadística y Geografía https://www.beta.inegi.org.mx/app/biblioteca/ficha.html?upc=702825216672 AYUNTAMIENTO DE HERMOSILLO https://www.hermosillo.gob.mx/portaltransparencia/

Tabla 4.1. Fuente de Datos

En esta sección se identificaron las fuentes de información externas e internas y, se realiza la selección del subconjunto de datos necesarios, los cuales se presentan a continuación:

A. Datos Climatológicos

Hay diversos lugares para consultar los datos climatológicos. El Servicio Nacional Meteorológico (SNM), organismo encargado de proporcionar información sobre el estado del tiempo a escala nacional y local en nuestro país, cuenta con diversas redes para bases de datos con estaciones meteorológicas. Los datos necesarios para esta investigación son los datos metrológicos de las variables relevantes para los años comprendidos entre el 2014 y 2017; en las estaciones registradas en la base de datos climatológica nacional (Sistema CLICOM). La ciudad de Hermosillo cuenta con 4 estaciones dentro de la ciudad, entre las cuales, solamente una presenta datos útiles para la investigación (figura 4.1), la estación 26139–HERMOSILLO II (DGE), ya que tiene datos comprendidos desde 1973 hasta 2016, sin embargo, deja una brecha desde octubre del 2016 a la actualidad, por lo que se tuvo que buscar otra red de base de datos que completara la información faltante. Dichos datos se obtuvieron de las Estaciones Meteorológicas Automáticas (EMAS) (figura 4.2), donde la información descargable está disponible solamente por 24 horas, una semana o 90 días previos a la fecha actual.

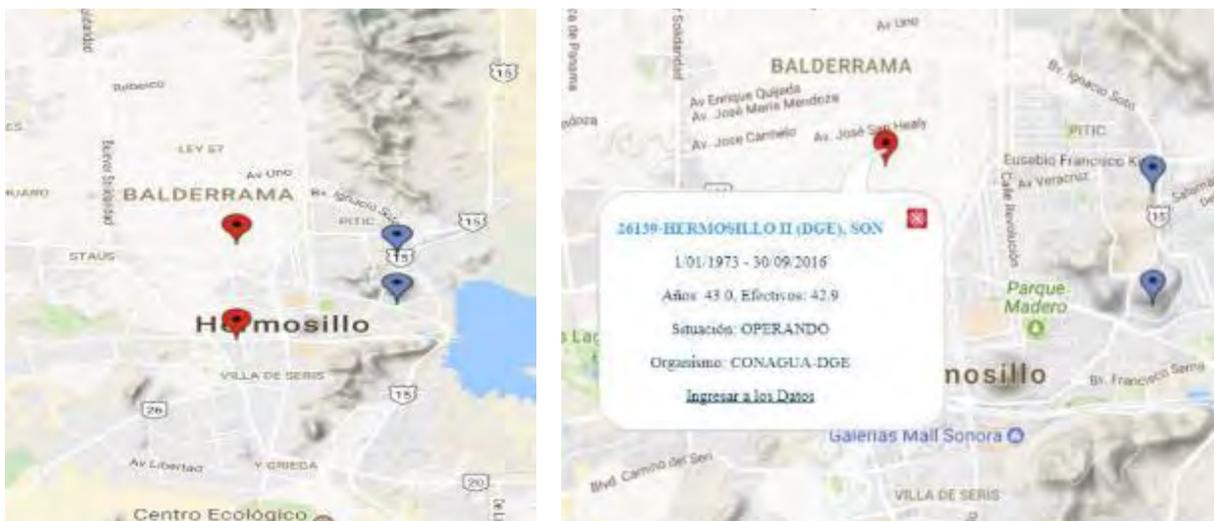


Figura 4.1. Estaciones Meteorológicas en Sonora registradas en CLICOM (Datos del mapa: Google, INEGI)



Figura 4.2 Estaciones Meteorológicas Automáticas (EMA)

Aunado a las estaciones necesarias para la investigación, se analizó la frecuencia de las variables medidas y su comportamiento en el periodo de análisis, lo que permitió determinar que las variables de precipitación (P), temperatura máxima (Tmax), temperatura mínima (Tmin) y temperatura promedio (Tprom) serían las variables de influencia de los datos climatológicos, por contar con un alto porcentaje de registros completos.

B. Datos Epidemiológicos

Los datos epidemiológicos analizados son suministrados por parte de una Institución Pública, en la cual, los derechohabientes afiliados representan el 10% de la población total. Los datos obtenidos de la Secretaría de Salud son datos históricos a nivel nacional.

4.2. Fase II. Construcción del modelo

En esta fase se encuentran los elementos de georreferenciación y minería de datos, primeramente, se realiza una exploración inicial de los datos, donde se integran y transforman los conjuntos de datos de las variables relevantes; obteniendo medidas relacionadas a los límites, cantidad de filas, valores perdidos y tipos de datos, y así, preparar el repositorio a ser utilizado por las herramientas de georreferenciación y las técnicas de minería de datos.

Paso 1. Preprocesamiento de datos

Una vez establecidas las estaciones meteorológicas que contienen la información necesaria se procedió a descargarla en formato .CSV y realizar un primer análisis con la herramienta Excel en las escalas mensual y anual, con el propósito de preparar los datos de entrada, por proceder de fuentes heterogéneas y no tener el formato adecuado, además de contener ruido.

Algunos datos fueron descartados durante este primer paso, por ejemplo, en los datos de rinofaringitis, la cantidad de datos antes del preprocesamiento es de 90,851 datos, los cuales, se comparan las colonias registradas a los pacientes, con la lista de colonias reconocidas por el H. Ayuntamiento de Hermosillo (Anexo B), considerando únicamente los datos con colonias registradas por dicha institución y los datos con registros completos, queda un total de 85,964 datos a utilizar.

Los datos que se descartaron, al considerar el registro de la colonia, no fueron utilizados con las herramientas de GIS, sin embargo, sí fueron considerados en los análisis relacionados a los grupos de edad, sexo, organismo y tipo de derechohabiente.

A. Datos Climatológicos

Los datos climatológicos que se utilizaron para crear un repositorio con todo el histórico de la información son:

- Precipitación (Pmm)

La figura 4.3 muestra la distribución y variabilidad mensual de la precipitación en la ciudad de Hermosillo. Como se puede observar, la temporada de lluvias se presenta entre junio y septiembre, mientras que la temporada de sequía es de octubre a mayo. Los meses con más lluvia son agosto y septiembre, con valores promedio entre 102 mm y 113 mm, y el mes con menor variabilidad es el mes de agosto. El mes con menor cantidad de lluvia es mayo con un valor promedio de 1 mm.

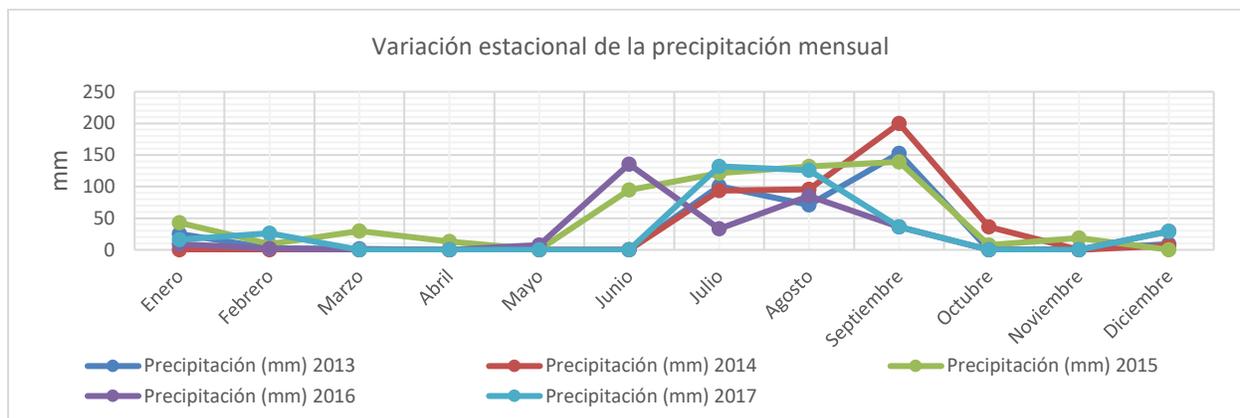


Figura 4.3 Variación mensual de la precipitación entre los años 2014 y 2017 en la ciudad de Hermosillo, Sonora, México. Fuente: Base de Datos Climatológica Nacional (CLICOM).

➤ Temperatura Promedio (T_{prom})

En la figura 4.4 se presenta la variabilidad mensual de la temperatura promedio, se observa un incremento gradual de temperatura de enero a junio, posteriormente se mantiene ligeramente constante en torno a los 33°C, para en septiembre comenzar a descender hasta alcanzar las temperaturas más bajas en enero. La mayor temperatura promedio registrada es el durante el 2014, en el mes de junio, con alrededor de 35°C.

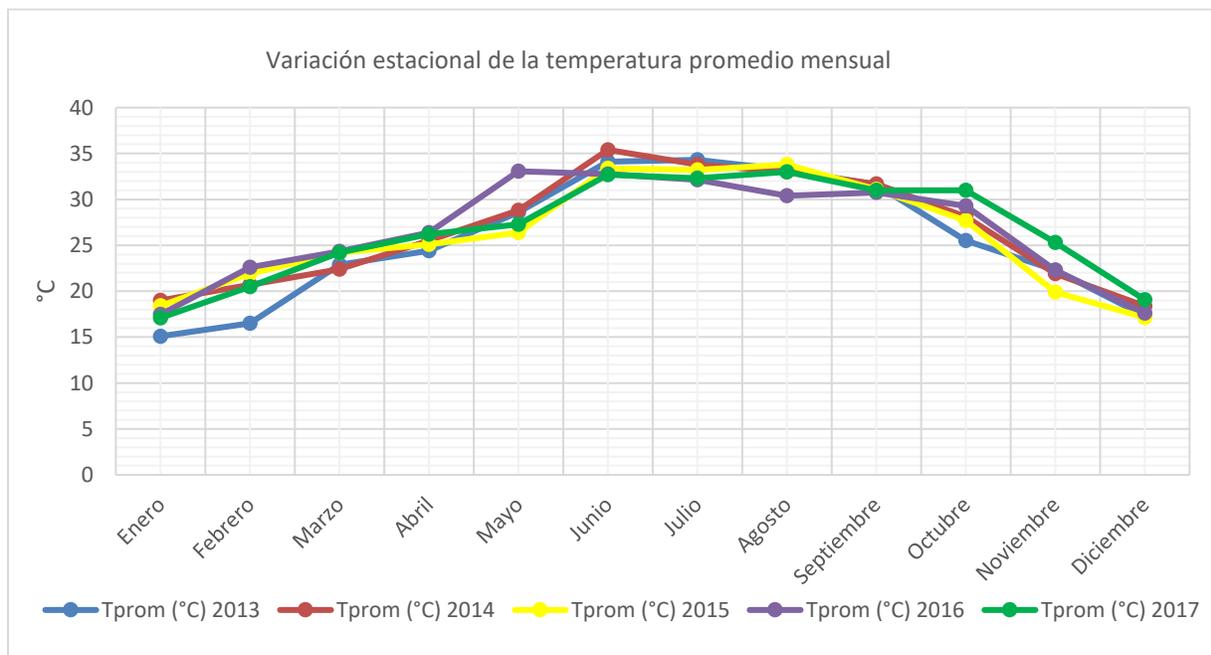


Figura 4.4. Variación estacional de la temperatura promedio mensual entre los años 2014 y 2017 en la ciudad de Hermosillo, Sonora, México. Fuente: Base de Datos Climatológica Nacional (CLICOM).

➤ Temperatura Máxima (Tmax)

En la figura 4.5 se observa un cambio similar de temperatura durante los años 2014, 2015 y 2017, un aumento gradual de enero a julio, para posteriormente comenzar a disminuir de agosto a diciembre, mientras que, durante el año 2014, el mes de junio alcanzó una temperatura de 49.5°C, siendo la temperatura más alta registrada, después de la cual comienza a descender. El año 2016 tuvo un comportamiento oscilante, totalmente distinto a los otros años, presenta una disminución inusual de alrededor de 5°C de temperatura, en los meses de junio a septiembre, con respecto a la media de los otros años.

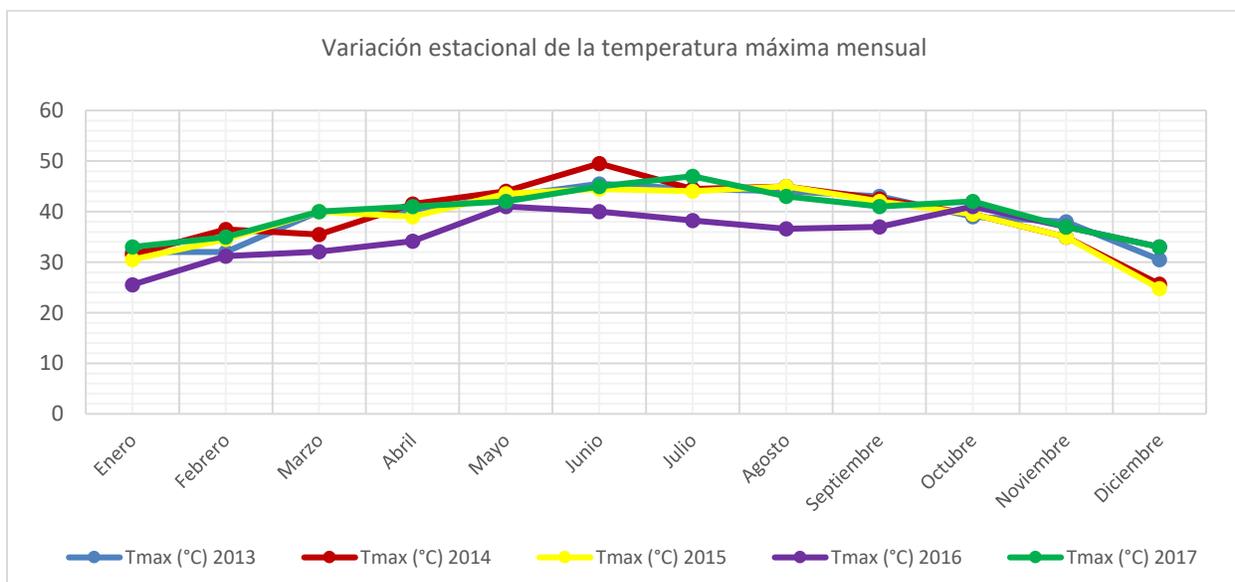


Figura 4.5. Variación estacional de la temperatura máxima mensual entre los años 2014 y 2017 en la ciudad de Hermosillo, Sonora, México. Fuente: Base de Datos Climatológica Nacional (CLICOM).

➤ Temperatura Mínima (Tmin)

Con respecto a la variabilidad mensual de las temperaturas mínimas registradas, se observa en la figura 4.6 que, de enero a junio, la temperatura mínima tiene un incremento constante alcanzando, por lo general, en julio las más altas temperaturas con oscilaciones entre los 23 y 26 °C, permaneciendo constante durante los meses de agosto y septiembre en temperaturas que varían entre los 18 y 24 °C. Durante los meses de octubre, noviembre y diciembre la temperatura desciende notablemente hasta alcanzar las temperaturas más frías del año entre los 0 y 4°C.

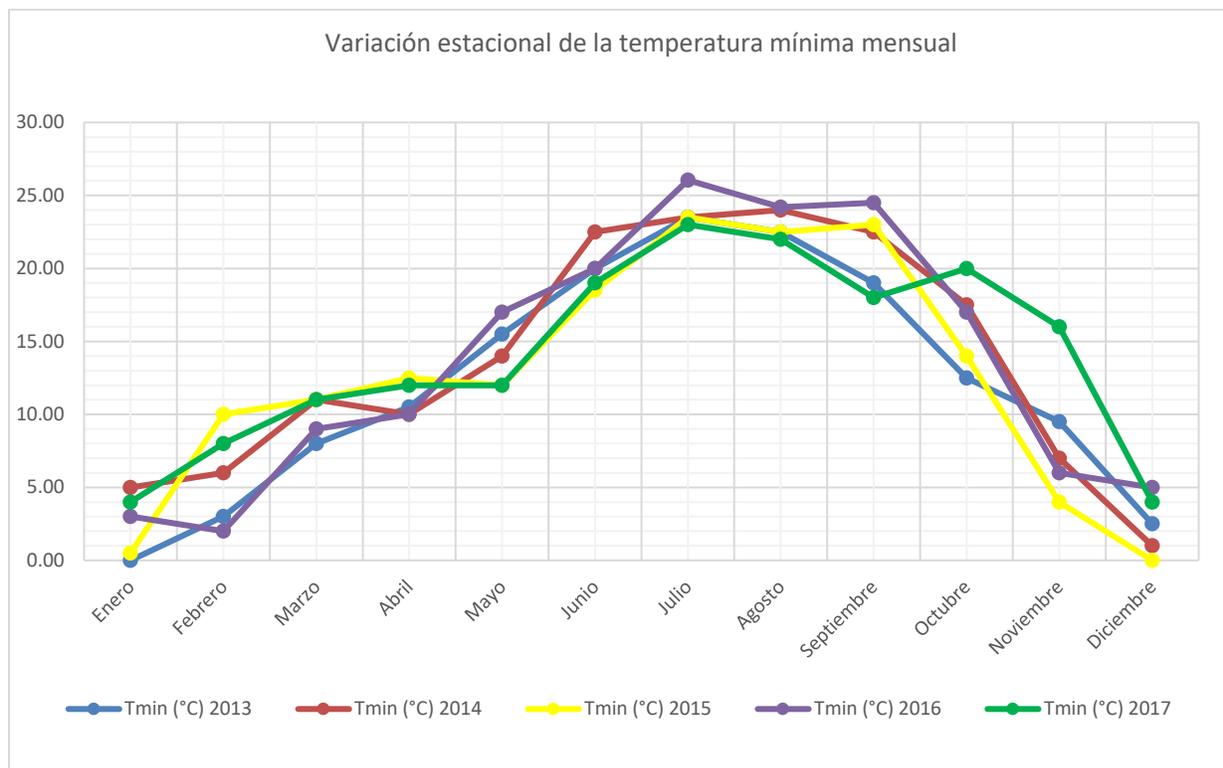


Figura 4.6. Variación estacional de la temperatura mínima mensual entre los años 2014 y 2017 en la ciudad de Hermosillo, Sonora, México. Fuente: Base de Datos Climatológica Nacional (CLICOM).

B. Datos Epidemiológicos

En la actualidad, la población de derechohabientes en Hermosillo es alrededor de 83,000 personas, los datos epidemiológicos con los que se cuentan son de los derechohabientes diagnosticados con las siguientes enfermedades: Dengue, Chikungunya, Rickettsiosis, Diarrea Aguda y Rinofaringitis. Dichas enfermedades se analizaron y utilizaron para la creación de un repositorio, y se les aplicó un primer análisis descriptivo obteniendo la siguiente información:

- **Dengue**

En la figura 4.7 se muestra como el año 2014, a pesar de no tener registros durante los primeros cinco meses del año, registro 3314 episodios siendo significativamente más dominante que el año 2015, con 2656 casos. La mayor incidencia se encontró en el mes de noviembre con 2111 (63.70%) para el 2014 y 1108 (41.72%) para el 2015, los años posteriores presentan un descenso de alrededor del 95%.

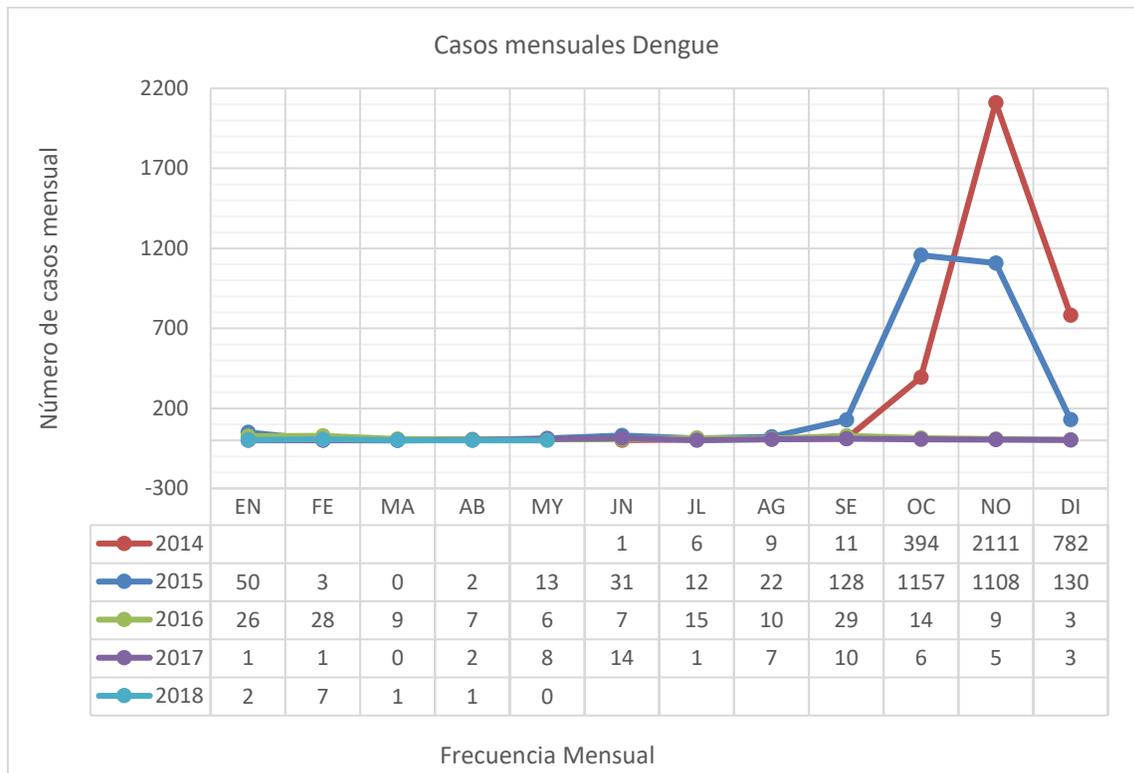


Figura 4.7. Comportamiento del número total de casos de dengue registrados entre los años 2014 y 2017 en el municipio de Hermosillo.

A partir del año 2016 se presenta una disminución de incidencias en casos de dengue en el municipio de Hermosillo. Una de las causas probables es la implementación de campañas de salud y fumigación. Otra de las causas posibles es que la población afectada en años anteriores adquiere inmunidad para el mismo serotipo (1, 2, 3, 4) pero sigue expuesta para el resto. Existe una disminución del 94% en casos de dengue del año 2015 al 2016.

En la figura 4.8 se observa, una intersección geométrica entre las capas de los diagnósticos por dengue anuales, la cual da por resultado, que la colonia Emiliano Zapata (colonia en azul claro), es la única que mantiene presente en los años analizados, otro factor interesante es la forma en cómo se encuentra rodeada por colonias con un alto índice de diagnósticos (figura 4.9).



Figura 4.8. Capas superpuestas de las colonias que presentaron diagnósticos durante el 2014, 2015, 2016, 2017 y 2018

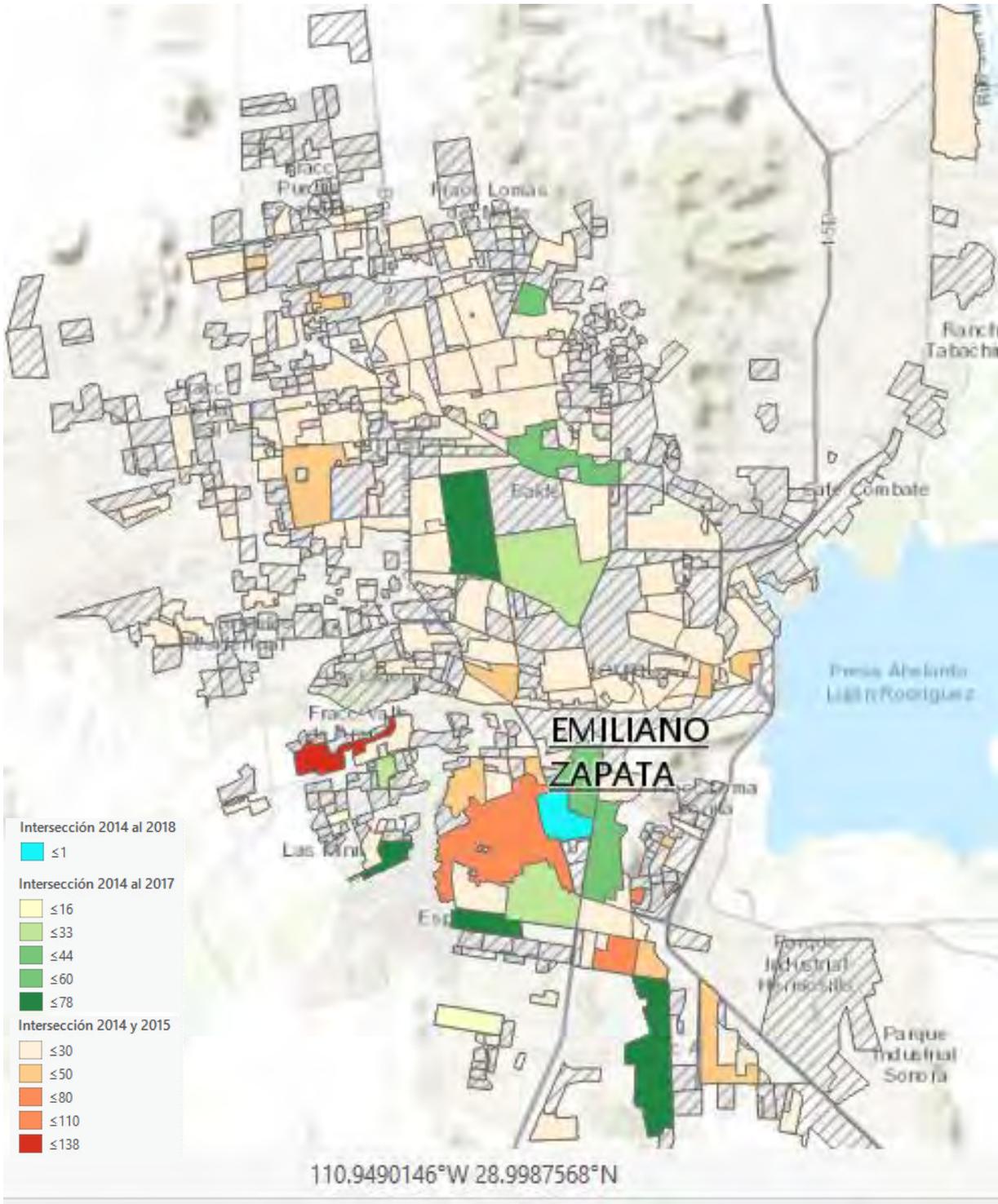


Figura 4.9 Intersección geométrica de las colonias con diagnósticos en los años 2014, 2015, 2016, 2017 y 2018.

- **Chikungunya**

Prosperó en el invierno del 2015, presentando tres colonias críticas en las cuales persistió la epidemia hasta la erradicación. Siendo estas El Mariachi, López Portillo y Margarita Sección Olivos. Todos los pacientes que fueron diagnosticados con esta enfermedad entre los años 2014 y 2017, sucedieron al norte y centro de la ciudad (figura 4.10), siendo la única enfermedad que no tiene presencia en la zona sur la ciudad.

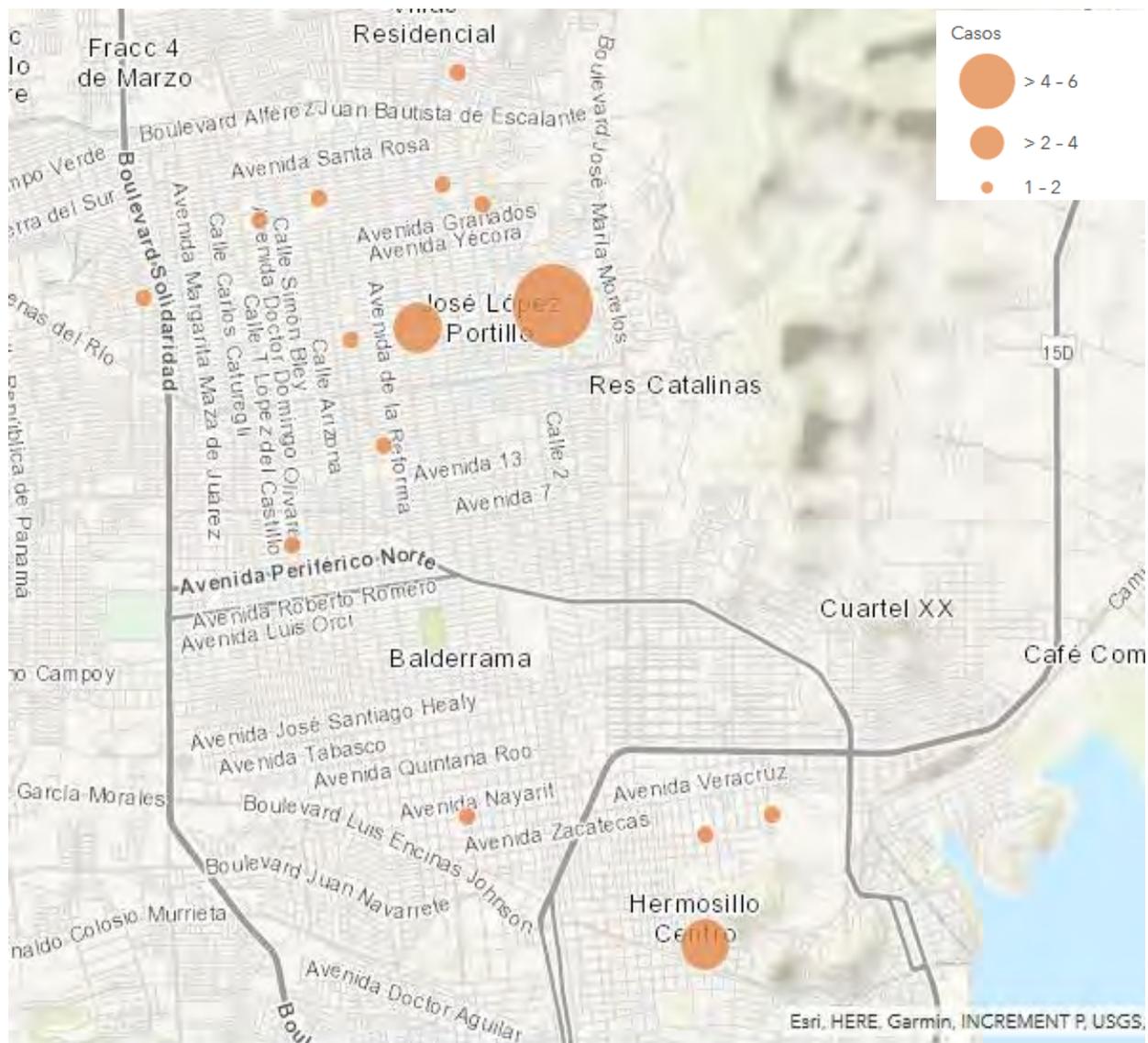


Figura 4.10 Mapa de los casos diagnosticados con Chikungunya entre los años 2014 y 2017

- **Diarrea Aguda**

Al comparar los casos del 2016 con los casos del 2017 (figura 4.12), se observa una disminución de la incidencia, lo cual puede estar asociado al cumplimiento satisfactorio del Programa Sectorial de Salud 2016-2021 y la mejora en la iniciativa "Caminando a la excelencia".

A pesar de que en el año 2017 se registraran mayores temperaturas máximas (figura 4.5), se mantuvieron bajo control los casos de diarrea aguda y con deshidratación, obteniendo una incidencia del 16% menor al año anterior.

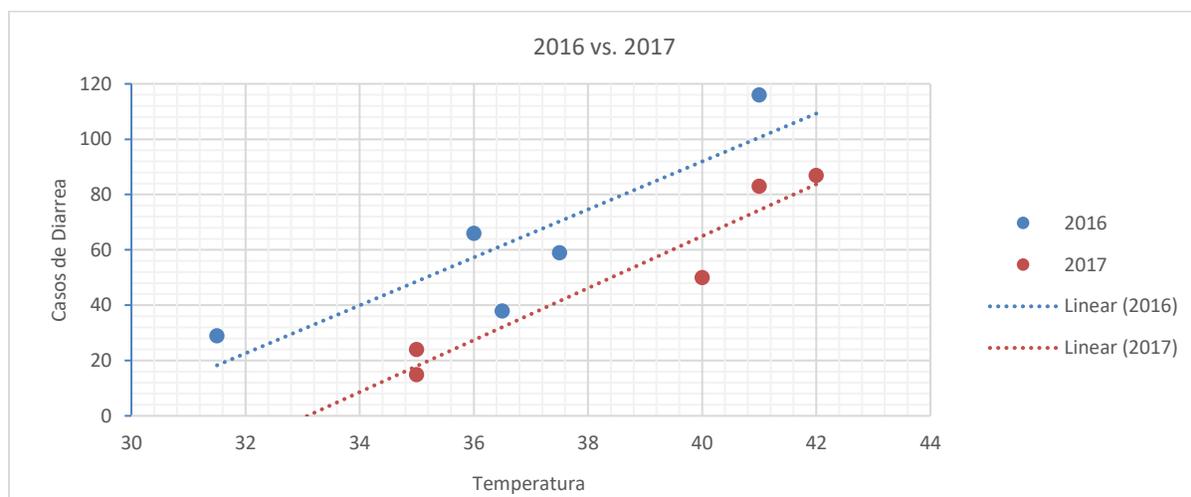


Figura 4.12. Comparación entre los años 2016 y 2017 con respecto a la temperatura

- **Rinofaringitis**

El análisis general del panorama epidemiológico actual de la rinofaringitis en el municipio de Hermosillo, analizando los 70,564 datos comprendidos entre el 01 de enero del 2014 al 31 de enero del 2018. Posteriormente se realiza un análisis más detallado para poder establecer si existen trabajadores de una misma dependencia que se enferman en períodos similares, si la rinofaringitis viene acompañada de otras enfermedades, si es recurrente en algunos pacientes y poder localizar a los pacientes que tienen más consultas al año registradas, así como para poder encontrar si hay períodos en los que hay más rinofaringitis en ciertas áreas de la ciudad.

➤ Análisis General

El análisis de los datos muestra un comportamiento similar a través de los años, de agosto a enero, los casos tienen a aumentar progresivamente mientras que en febrero comienzan a disminuir, hasta llegar a julio, donde la incidencia de rinofaringitis es muy escasa (figura 4.13).

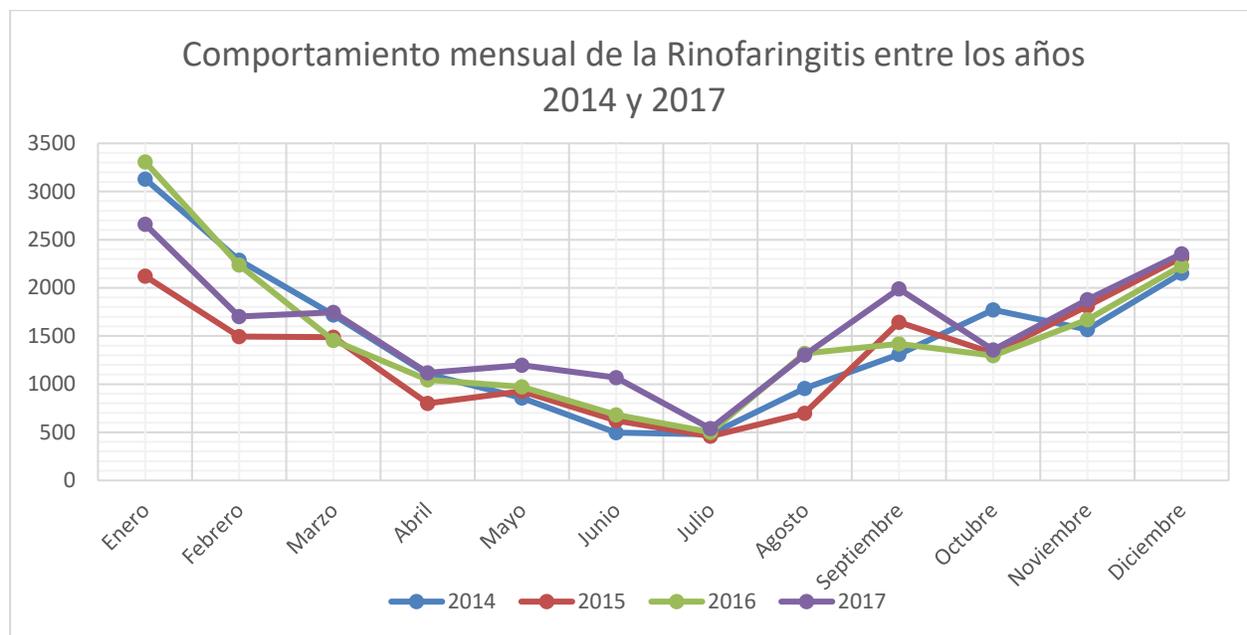


Figura 4.13. Comportamiento mensual de la rinofaringitis entre los años 2014 y 2017 en el municipio de Hermosillo, Sonora

- ¿Quiénes fueron los más afectados por rinofaringitis?

El mayor número de diagnósticos con esta enfermedad recae en los niños que tienen entre 2 y 9 años, entre la edad de 10 y 19 años los casos disminuyen, para posteriormente, de los 20 a 65 años o más, mantenerse entre los 2000 y 4000 casos (figura 4.14). Como consecuencia a esto, el tipo de derechohabiente más afectado son los hijos de los trabajadores, seguido por los trabajadores y en tercer lugar las esposas de los trabajadores (figura 4.15). En el año 2017, se registraron 3676 diagnósticos de rinofaringitis en trabajadores, quienes pertenecen a 55 dependencias en total, de las cuales, en seis (Gobierno del Estado, H. Ayuntamiento, ISSSTESON, Magisterio, Servicios de Salud de Sonora, UNISON) se concentra el 80% de las incidencias y en tan

solo dos de ellas el 53%, siendo estas el Gobierno del Estado (36.48%) y el H. Ayuntamiento (17.22%).

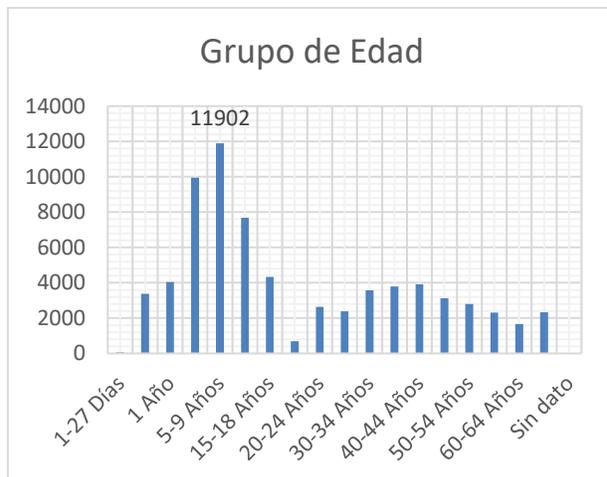


Figura 4.14. Casos de rinofaringitis por grupo de edad registrados entre los años 2014 y 2018 en el municipio de Hermosillo, Sonora

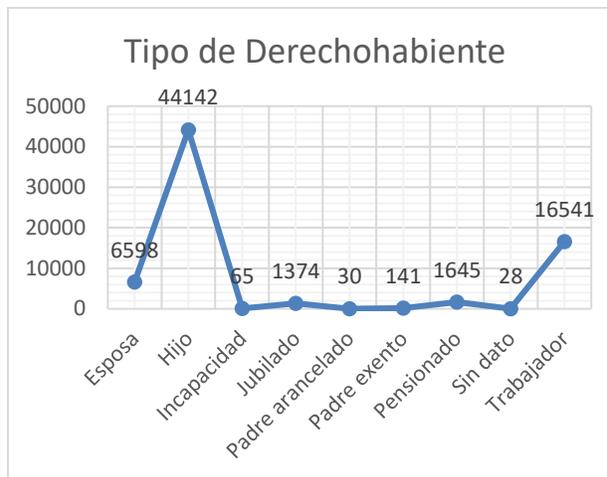


Figura 4.15. Casos de rinofaringitis por tipo de derechohabiente registrados entre los años 2014 y 2018 en el municipio de Hermosillo, Sonora

Analizando las seis dependencias principales (figura 4.16), se obtuvo un comportamiento similar entre ellas (figura 4.17 y 4.18), una alta incidencia durante enero, siendo el mes con mayor número de casos registrados, para bajar gradualmente hasta julio, posteriormente sube en agosto alcanzando un pico en septiembre, disminuyendo brevemente durante el mes de octubre para comenzar a ascender nuevamente; resultando afín al comportamiento mensual que tuvo la enfermedad (figura 4.13).

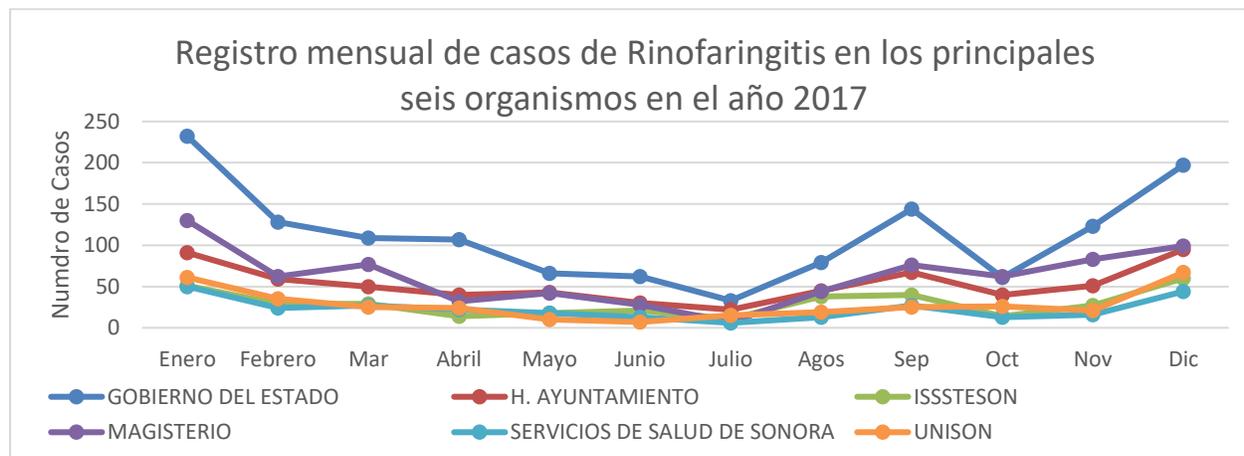


Figura 4.16. Registro de diagnóstico mensual durante el 2017 de rinofaringitis en trabajadores de las seis dependencias más significativas

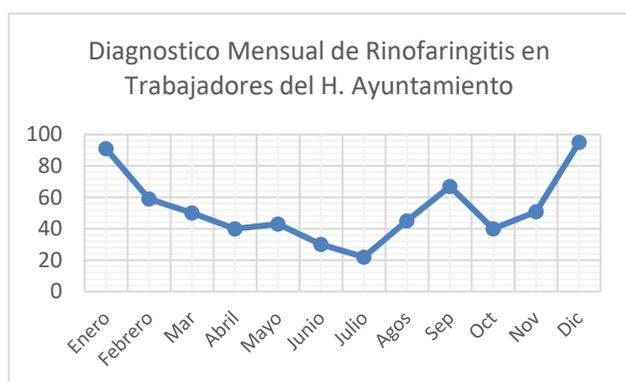


Figura 4.17. Diagnostico mensual de rinofaringitis en trabajadores del H. Ayuntamiento durante el año 2017

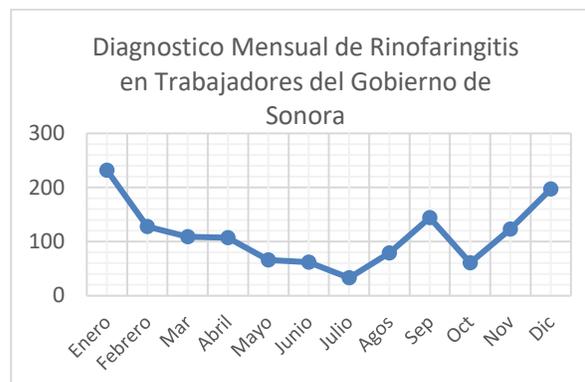


Figura 4.18. Diagnostico mensual de rinofaringitis en trabajadores del Gobierno de Sonora durante el año 2017

- Enfermedades Recurrentes

De los pacientes que fueron diagnosticados con rinofaringitis, el 35.83% también sufrió de Sinusitis Aguda, el 35.02% de faringitis, el 100% de los pacientes diagnosticados con enfermedad de las vías aéreas debido a otros polvos orgánicos específicos también fueron diagnosticados con rinofaringitis, al igual que el 37.14% de las personas con algún tipo de diagnóstico relacionado con las vías respiratorias superiores, como lo es la infección aguda de las vías respiratorias superiores (IRA) o reacción de hipersensibilidad de las vías respiratorias superiores.

Según la tabla 4.2 el 18.60% de los diagnósticos registrados en el año 2017 pertenece a pacientes recurrentes, principalmente los hijos de los trabajadores, con 13.03%, seguido por los mismos trabajadores con un 3.53%.

Derechohabiente	2 visitas	≥3 y <5 visitas	≥5 visitas	Total	%
Esposa	216	50	8	274	1.45%
Hijo	1614	720	129	2463	13.03%
Incapacidad	0	0	0	0	0.00%
Jubilado	33	6	0	39	0.21%
Padre arancelado	1	0	0	1	0.01%
Padre exento	2	0	0	2	0.01%
Pensionado	61	7	1	69	0.36%
Trabajador	540	117	11	668	3.53%
Total general	2467	900	149	3516	18.60%
%	13.05%	4.76%	0.79%	18.60%	

Tabla 4.2. Derechohabientes que presentaron recurrencia en diagnósticos de rinofaringitis durante el 2017 en el municipio de Hermosillo, Sonora

Se encontró que la enfermedad si es recurrente en trabajadores de distintas dependencias, 540 casos de pacientes registraron dos visitas durante el año, 117 con tres o cuatro visitas y once con cinco o más visitas; las cuales sucedieron con mayor frecuencia entre los meses de agosto y enero, siendo diciembre y enero los meses con mayores afectaciones.

- Georreferenciación de la rinofaringitis

En la figura 4.19 se observan las colonias que presentaron registros de rinofaringitis durante el año 2017, la enfermedad permanece más constante en zonas cercanas a cerros, recorriendo tanto el norte como el sur de la ciudad; al analizar los datos con mayor profundidad, se obtuvo que en tan solo el 6.4%, del total de las colonias con diagnósticos registrados, se presenta el 38.53% de casos de rinofaringitis que hubo durante el año (figura 4.20).

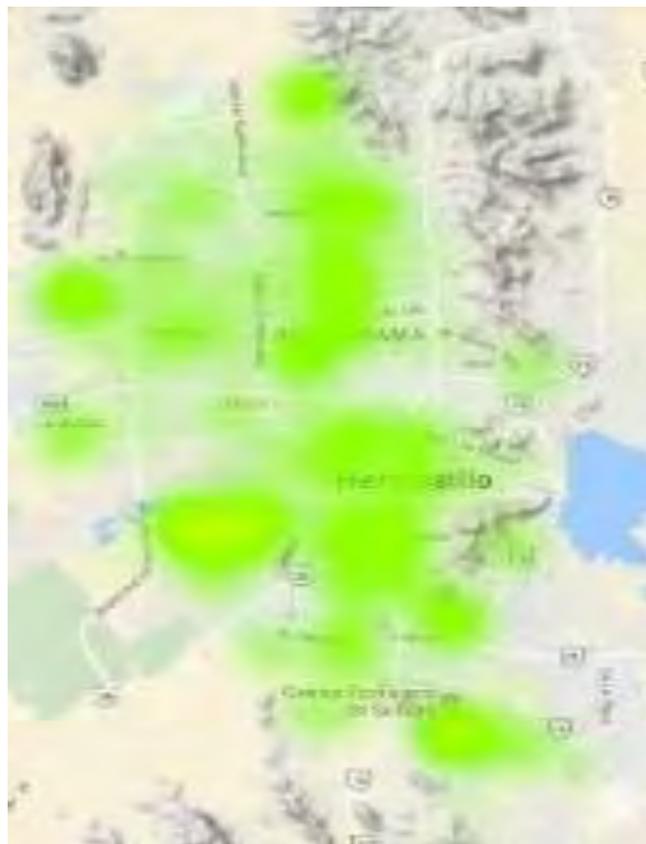


Figura 4.19. Colonias que registraron diagnósticos de rinofaringitis durante el año 2017 en la ciudad de Hermosillo, Sonora

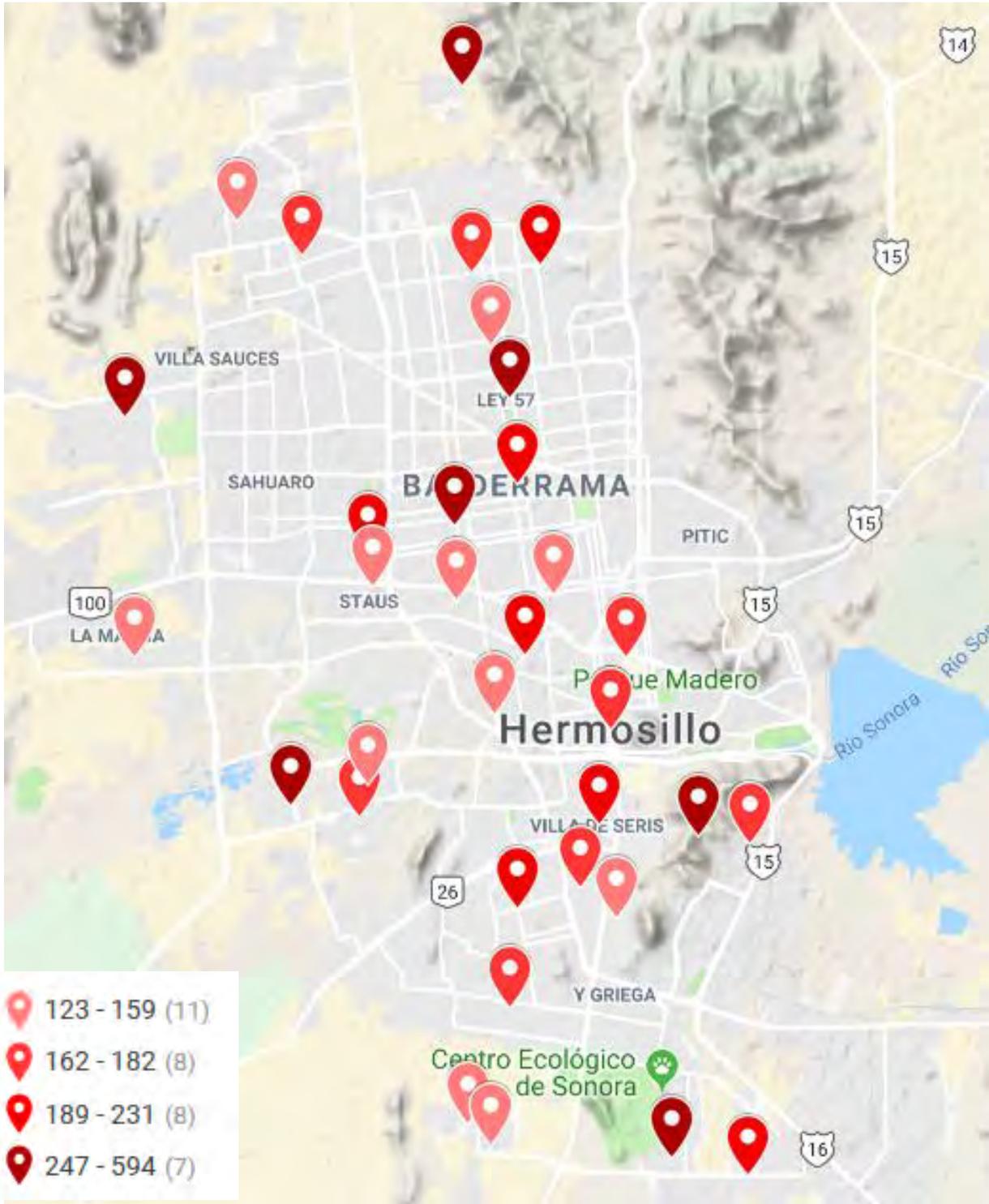


Figura 4.20. Colonias con mayor incidencia de rinofaringitis durante el 2017 en la ciudad de Hermosillo, Sonora

Al comparar los dos meses con mayor cantidad de registros (enero y diciembre) durante el año 2017, se obtiene que, a pesar de que la enfermedad se encuentra dispersa en casi toda la ciudad de Hermosillo, la mayor cantidad de incidencia sucede en la zona centro y sur (figura 4.21 y 4.22).



Figura 4.21. Colonias con mayor incidencia durante el mes de enero del 2017 en la ciudad de Hermosillo, Sonora

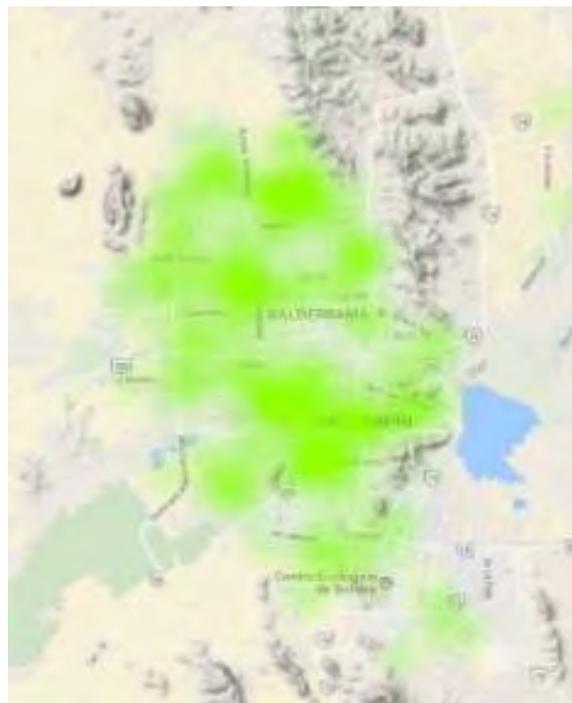


Figura 4.22. Colonias con mayor incidencia durante el mes de diciembre del 2017 en la ciudad de Hermosillo, Sonora

También se encontró que en dichos meses (enero y diciembre), reinciden siete colonias (figura 4.23) con el mayor número de casos, las cuales se encontraron presentes en la mayoría del año (tabla 4.3)

Colonia	Enero 2017	Diciembre 2017
Altares	91	56
Villa Bonita	70	68
Puerta Real	53	37
Olivares	49	61
Villa Sonora	33	39
Palo Verde Sur	33	30
Ley 57	36	48

Tabla 4.3. Colonias con mayor cantidad de diagnósticos de rinofaringitis en enero y diciembre del año 2017



Figura 4.23. Colonias con mayor cantidad de diagnósticos tanto en enero como en diciembre del 2017 en la ciudad de Hermosillo, Sonora

La tabla 4.4 muestra los trabajadores con mayor número de diagnósticos de rinofaringitis durante el año 2017.

≥ 5				
H. Ayuntamiento	Consultas al año	Colonia	Gpo. Edad	Sexo
xxxxx01	12	COLONIA OLIVARES	60-64 Años	Masculino
xxxxx501	6	COLONIA HACIENDA DE LA FLOR	55-59 Años	Masculino
xxxxx501	5	COLONIA LAS VILLAS	50-54 Años	Masculino
Gobierno del estado				
xxxx701	8	COLONIA REVOLUCION	65 Años o más	Masculino
xxxx601	6	COLONIA VILA DE SERIS	65 Años o más	Masculino
xxxxx201	5	COLONIA PASEO DE LAS MISIONES	40-44 Años	Masculino
xxxx701	5	COLONIA LA RIOJA	45-49 Años	Masculino
Magisterio				
xxxx201	5	COLONIA NACAMERI	50-54 Años	Femenino
ISSSTESON				
xxxxx301	5	FRACC. LUIS DONALDO COLOSIO	25-29 Años	Femenino
UES				
xxxx901	5	COLONIA SAHUARO	55-59 Años	Masculino
Servicios de S				
xxxxx001	5	FRACCIONAMIENTO LAS LOMAS	35-39 Años	Femenino

Tabla 4.4. Derechohabientes de tipo trabajador con mayor recurrencia de diagnóstico

• **Comorbilidad**

Mediante el uso de tablas de contingencia sencillas, se cuantificó la comorbilidad (figura 4.24), en pares, de las enfermedades analizadas y se obtuvieron los siguientes resultados:

Diccionario:

ATRIBUTO	DIAGNÓSTICO
D1	RINOFARINGITIS
D2	FARINGITIS AGUDA
D3	DIARREA
D4	DENGUE
D5	RICKETTSIOSIS
D6	CHIKUNGUNYA
D7	OBESIDAD

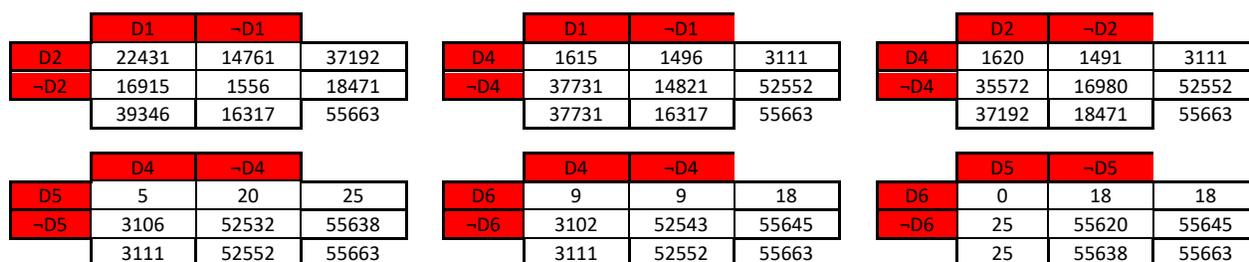


Figura 4.24. Resultados obtenidos de Comorbilidad con mayor significancia entre la Rinofaringitis, Faringitis Aguda, Diarrea, Dengue, Rickettsiosis, Chikungunya y Obesidad.

Con el propósito de visualizar los resultados obtenidos, se elaboró la siguiente gráfica:

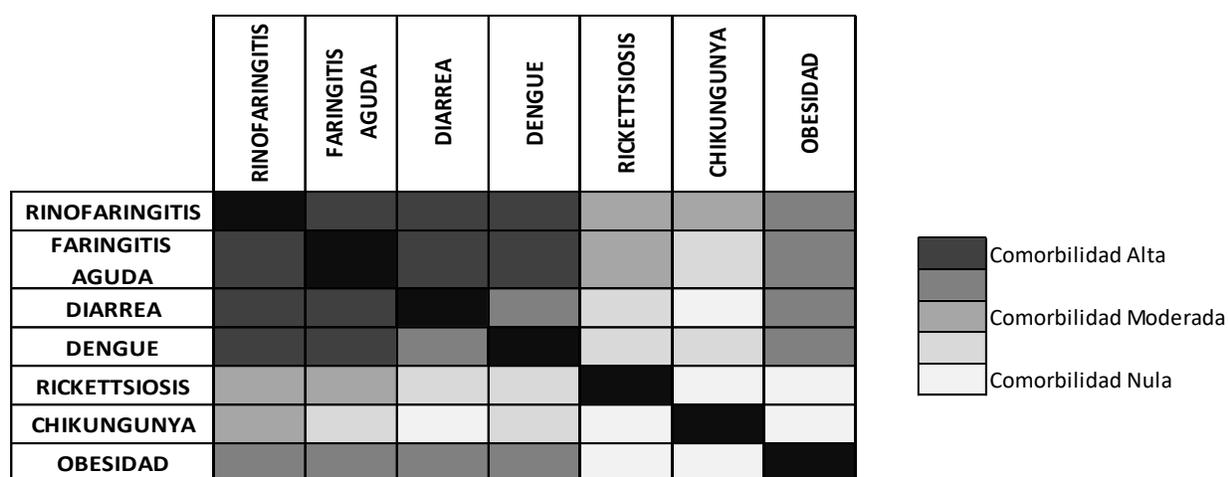


Figura 4.25. Mapa de Calor según la comorbilidad encontrada entre las distintas enfermedades

- **Caracterización de Enfermedades**

Después de realizar estudios observacionales, se pretende explicar y describir sobre la ocurrencia, magnitud, distribución y progresión de la salud y de las causas de comorbilidad, esto con el propósito de dar respuesta a preguntas como ¿A quién?, ¿Cuándo? y ¿Dónde?; la caracterización de las enfermedades se realizó, primeramente, analizando las variables demográficas como grupo de edad y sexo, donde la relación entre estas variables y la enfermedad, se suele describir mediante tasas específicas de frecuencia. En la tabla 4.5 se muestran las diferentes características de exposición a un posible brote de enfermedades.

Diagnóstico	Gpo de Edad	Sexo	Tipo Derechohabiente	Organismo	Tprom	Tmax	Tmin	Precipitación	Estación
Dengue	10-18 años 23%	Femenino 54%	Hijo 34%	Gobierno del estado 29%	31°C mes anterior	38°C mes anterior	24°C mes anterior	116 mm mes anterior	Otoño 92%
Chikungunya	30-34 años 23%	Femenino y Masculino 50%	Trabajador 90%	Gobierno del estado 40%	19°C mes anterior	26°C mes anterior	11°C mes anterior	10 mm mes anterior	Otoño 63%
Rickettsiosis	2-9 años 34%	Femenino 54%	Hijo 34%	Gobierno del estado 29%	32 °C	39 °C	26 °C	166 mm	Verano 45%
Diarrea	35-45 años 17%	Femenino 54%	Trabajador 50%	Gobierno del estado 31%	30 °C	38 °C	22 °C	10 mm	Primavera 40%
IVR*	1-9 años 33%	Femenino 58%	Hijo 57%	Gobierno del estado 31%	26 °C	33 °C	19 °C	1 mm	Invierno 31%

Tabla 4.5. Caracterización por enfermedad

*IVR = Infección de las Vías Respiratorias

Paso 2. Aplicación de técnicas de minería de datos

Después del análisis previo de los datos, se observó que es necesario evaluar similitudes entre los datos para encontrar subgrupos de datos, con el propósito de que, dentro de dichos subgrupos, los pacientes tengan características similares.

Razón por la cual, el agrupamiento, es la técnica de minería de datos que se ajusta al planteamiento del problema, ya que se crean los subgrupos de pacientes basándose en sus características clínicas asociadas, patrones espaciales y temporales.

La creación del repositorio, al cual se les aplicó la técnica de agrupamiento, fue por medio de Colaboratory, los datos se encuentran separados en Hojas de Excel por tipos de enfermedad. En esta libreta se unieron todos los registros en una gran tabla, para poder procesarlos al realizar una matriz de diagnósticos y obtener clúster vía K-means sobre los mismos, como se muestra a continuación:

1. **Acceso a Google Drive:** Es necesario acceder con nuestras credenciales a Google Drive.
2. **Listar:** Se obtiene un listado del directorio donde están los datos, al igual que el nombre e ID de cada archivo.
3. **Cargar hojas de cálculo:** Se cargan las hojas de cálculo en una lista de dataframes (de pandas), para luego concatenar los datos en un solo dataframe maestro.
4. **Concatenar:** Se hace una sola tabla de datos con rinofaringitis, obesidad, diabetes, dengue, rickettsia y diarrea.
5. **Diccionarios de features (características):** Este paso es con el propósito de poder hacer las características en forma binaria, se consideraron como características el diagnóstico, grupo de edad, tipo de derechohabiente, género, estado civil y organismo.
6. **Lista de features binarias:** Esta sección consiste en formar las características.

7. **Agrupar por número de afiliación:** Se crea un dataframe con los números de veces que un Afiliado se ha enfermado, considerando solo a los que se han enfermado más de una vez.
8. **Agrupamiento:** Es la sección en la cual se obtienen los clústeres vía K-means
9. **Resultados:** A continuación, se presentan los resultados obtenidos.

Después de probar y comparar diferentes tamaños de núcleos, se obtiene que al igualar K a 3 (figura 4.32) se obtienen que los pacientes diagnosticados con Diabetes Mellitus, Diarrea con Deshidratación, Faringitis Aguda y Rinofaringitis Aguda, poseen características similares, si se aumenta el número de K, solo se obtienen subgrupos con estas mismas características (Anexo C).

```

cluster 0 Registros: 20132
=====
d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.05752036558712498
d18 DIABETES MELLITUS NO INSULINODEPENDIENTE,SIN MENCION DE COMPLICACION : 0.2211404728789986
d28 FARINGITIS AGUDA : 0.815666600437115
d56 RINOFARINGITIS AGUDA [RESFRIADO COMUN] : 0.9123783031988874

cluster 1 Registros: 4885
=====
d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.0065506653019447286
d4 DIABETES MELLITUS ASOCIADA CON DESNUTRICION,CON COMPLICACIONES CIRCULATORIAS PERIFERICAS : 0.006960081883316274
d18 DIABETES MELLITUS NO INSULINODEPENDIENTE,SIN MENCION DE COMPLICACION : 0.010849539406345957
d21 DIARREA CON DESHIDRATAACION : 0.04605936540429888
d28 FARINGITIS AGUDA : 1.0569089048106448
d56 RINOFARINGITIS AGUDA [RESFRIADO COMUN] : 4.310542476970317

cluster 2 Registros: 3936
=====
d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.001524390243902439
d3 DIABETES MELLITUS ASOCIADA CON DESNUTRICION,CON COMA : 0.0038109756097560975
d4 DIABETES MELLITUS ASOCIADA CON DESNUTRICION,CON COMPLICACIONES CIRCULATORIAS PERIFERICAS : 0.014735772357723578
d18 DIABETES MELLITUS NO INSULINODEPENDIENTE,SIN MENCION DE COMPLICACION : 0.033790650406504065
d21 DIARREA CON DESHIDRATAACION : 0.07240853658536585
d28 FARINGITIS AGUDA : 4.1763211382113825

```

Figura 4.26. Resultado del clúster con $k=3$

De acuerdo a los resultados obtenidos de pasar la matriz de diagnósticos por el método de agrupamiento (K-means clustering), se encuentra que el primer grupo (Clúster 0) consta de 20,132 registros (69.5% de la muestra) donde la mayoría de los pacientes fueron diagnosticados con Rinofaringitis o Faringitis aguda al menos una vez, y muy

infrecuentemente con Diabetes Mellitus. Cerca del 25% de los derechohabientes fue atendido por esta combinación de enfermedades durante enero del 2014 a mayo 2018.

El segundo grupo (Clúster 1), con 4,885 registros (16.9% de la muestra) revela pacientes que fueron diagnosticados más de 4 veces con Rinofaringitis en el periodo de estudio, faringitis aguda al menos 1 vez y muy pocas veces con diabetes mellitus con diarrea y deshidratación.

El tercer grupo (Clúster 2), con 3,936 registros (13.6% de la muestra) exhibe personas que asistieron al médico al menos 4 veces por Rinofaringitis, Diarrea y casos de Diabetes Mellitus con y sin mención de complicación.

Casi 5% de los derechohabientes fue atendido por más de 3 enfermedades en el periodo de evaluación, lo que sugiere que se podrían enfocar esfuerzos de medicina preventiva en este grupo de personas.

- **Clúster Jerárquico**

Al generar clústeres jerárquicos sobre la matriz con la caracterización por enfermedad, se obtuvieron, tanto con el software RStudio (figura 4.33) como con IBM SPSS Statistics (figura 4.34), dos agrupamientos como resultado, el primer clúster, es el grupo de las IVR, chikungunya y diarrea, el segundo clúster muestra una relación entre dengue y Rickettsiosis, donde el factor común que los asocia es la precipitación.

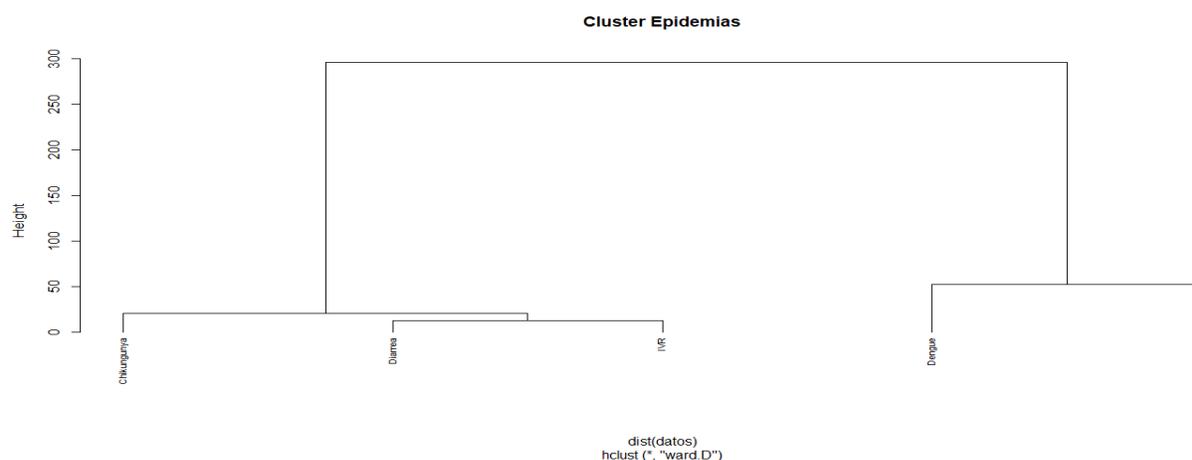


Figura 4.27. Dendograma obtenido en RStudio

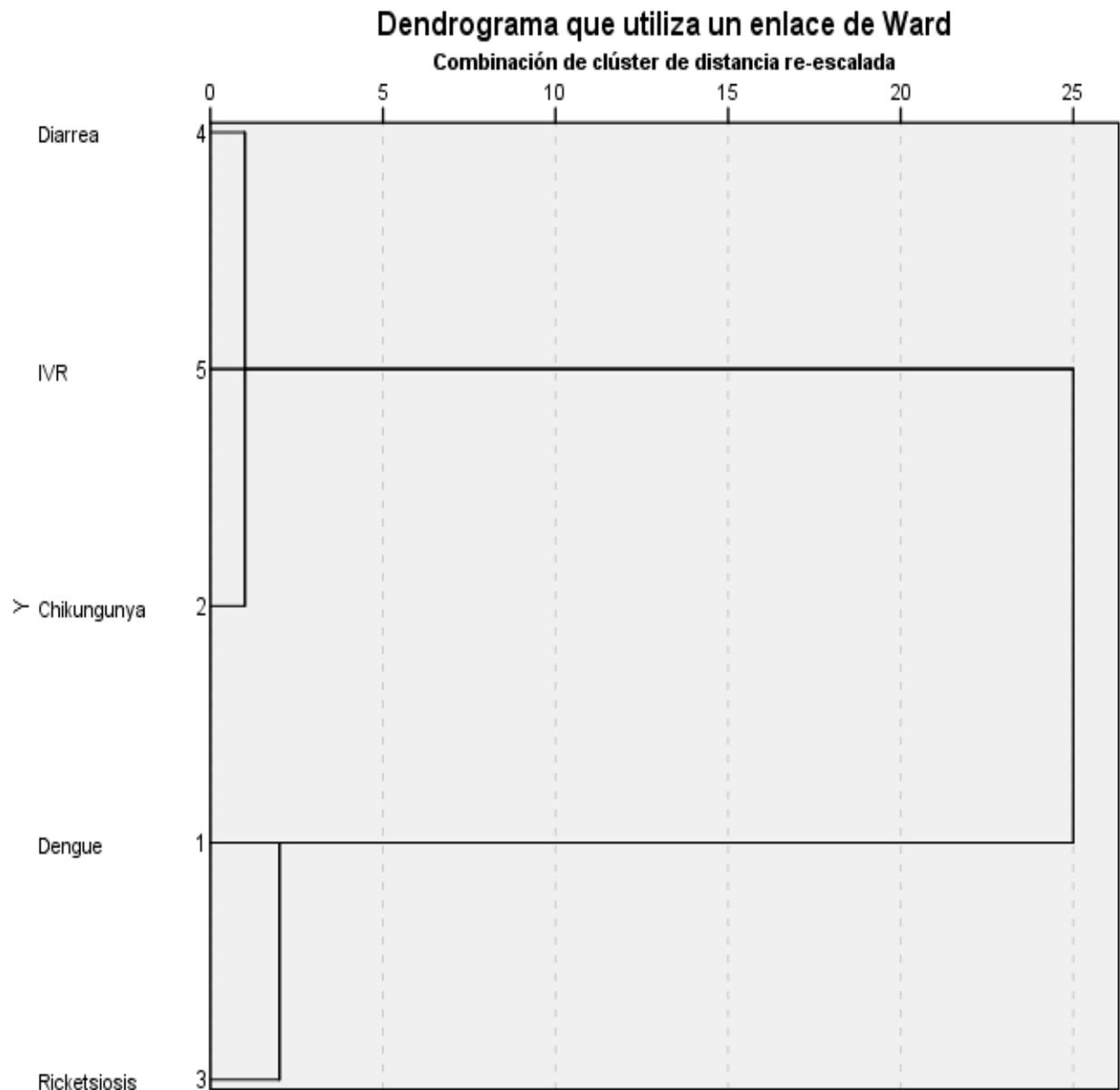


Figura 4.28. Dendrograma obtenido en SPSS

- Fase de pruebas y verificación de resultados

A continuación, se presentan los modelos de clasificación generados con la ayuda del software Weka y RapidMiner. Para la construcción del árbol de decisión se utilizó un set de datos de entrenamiento con 29,125 registros (60%), y para probarlo se utilizaron

19,417 registros (40%), con un margen de confianza cerca del 50% (figura 4.35Figura 4.29).

```

Classifier output

=== Evaluation on training set ===

Time taken to test model on training data: 0.22 seconds

=== Summary ===

Correctly Classified Instances      24227      49.9083 %
Incorrectly Classified Instances    24316      50.0917 %
Kappa statistic                    0.2478
Mean absolute error                 0.0363
Root mean squared error             0.1347
Relative absolute error             82.2726 %
Root relative squared error         90.7145 %
Total Number of Instances          48543
  
```

Figura 4.29. Resultados confiabilidad de los árboles de decisión

Los resultados con mayor relevancia obtenidos de la matriz son:

- 717 fueron diagnosticados con diarrea con deshidratación
- 1 Diarrea y gastroenteritis y diarrea con deshidratación
- 107 Enfermedad diarreica aguda y diarrea con deshidratación
- 5 Enfermedad diarreica aguda bacteriana EDA y diarrea con deshidratación
- 7 Enfermedad por virus chikungunya
- 1270 Fiebre del dengue (dengue clásico)
- 10 Fiebre del dengue hemorrágico y Fiebre del dengue (dengue clásico)
- 1 Fiebre maculosa Rickettsiosis transmitida por garrapatas
- 5005 Rinofaringitis aguda (resfriado común)
- 2326 Rinofaringitis aguda (resfriado común) y faringitis aguda
- 4357 Faringitis Aguda y Rinofaringitis aguda (resfriado común)
- 2302 Faringitis Aguda
- 420 Rinitis, Faringitis Crónicas y Rinofaringitis aguda (resfriado común)
- 193 Faringitis Bacteriana y Rinofaringitis aguda (resfriado común)
- 117 Laringofaringitis Aguda y Rinofaringitis aguda (resfriado común)
- 61 Laringofaringitis Aguda y Faringitis Aguda
- 31 Faringitis Crónica y Rinofaringitis aguda (resfriado común)
- 22 Faringitis Aguda Debida a otros microorganismos especificados y Rinofaringitis aguda (resfriado común)
- 58 Rinofaringitis Crónica y Faringitis Aguda

- 44 Faringitis Estreptococica y Rinofaringitis aguda (resfriado común)
- 14 Faringitis Estreptococica y Faringitis Aguda
- 16 Faringoamigdalitis Aguda y Rinofaringitis aguda (resfriado común)
- 9 Faringoamigdalitis y Faringitis Aguda
- 2 Rinolaringitis Aguda y Rinofaringitis aguda (resfriado común)
- 1 Rinolaringitis Aguda y Faringitis Aguda

En estos resultados se muestran predominantes las IVR, esto se debe a que la cantidad de registros, con estos diagnósticos en la base de datos, es del 89%. La figura 4.36 muestra como nodo raíz a la variable estación en el árbol obtenido en Weka, mientras que en la figura 4.37 se muestra la variable de precipitación (mm) y como nodo hijo a la temperatura promedio, las cuales, partiendo de la tabla 4.5, se resumen en la variable estación, por lo que, los dos árboles de decisión muestran el mismo resultado, solo que de maneras distintas.

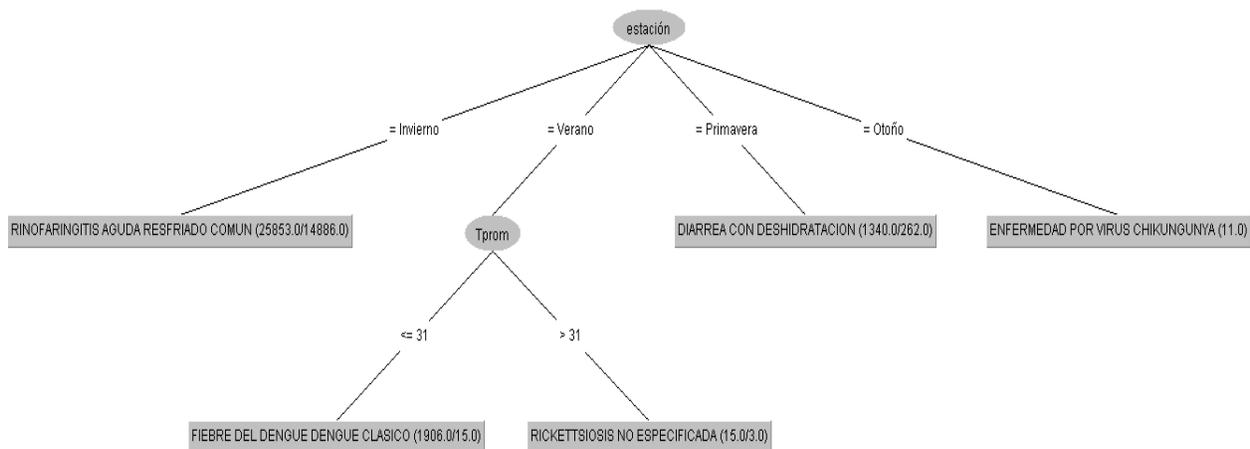


Figura 4.30. *Árbol de Decisión obtenido en el software Weka*



Figura 4.31. Árbol de Decisión Obtenido en el Software RapidMiner

En los resultados podemos encontrar una o varias reglas que se crean al relacionar diferentes atributos de las epidemias de rápida propagación, estas contienen probabilidades de ocurrencia y una importancia asociada para cada enfermedad, a su vez determinan los patrones de diagnóstico.

Las reglas identificadas permitirán a la Subdirección de Servicios Médicos conocer los diagnósticos con mayores incidencias, lo cual por un lado sirve para emitir alertas y preparar una pronta y eficaz respuesta al igual que orientar y fundamentar las acciones de salud, por otro lado, con el análisis de la información se llega a conocer aspectos desconocidos que serán aprovechados para favorecer las estrategias de medicina preventiva, logrando el bienestar social y la calidad de vida de la comunidad en general.

Los resultados encontrados pueden ser de utilidad en Instituciones de Salud Pública que se encuentren en distintos estados y/o municipios.

5. CONCLUSIONES, RECOMENDACIONES Y TRABAJOS FUTUROS

Esta sección comprende las conclusiones obtenidas a partir del trabajo de investigación y su correspondiente aplicación. Se mencionan recomendaciones, así como trabajos futuros de investigación que se pueden establecer con base en los resultados obtenidos en este trabajo de tesis.

5.1. Conclusiones

En este trabajo se trataron los problemas asociados a la detección de brotes epidemiológicos de rápida propagación, definiendo para ello una metodología de trabajo con la cual implementar mejoras en la toma de decisiones mediante el análisis de georreferenciación apoyado con técnicas de minería de datos. En este sentido se cumplieron todos los objetivos planteados, obteniéndose los resultados asociados a cada uno de ellos:

Se establecieron procedimientos para el análisis de datos y detección de un comportamiento atípico en la demanda de atención. El proceso de análisis y preparación de datos fue el más complicado, debido a la información disponible y su variabilidad, así se pasó de análisis anuales, a mensuales para ocasionalmente, según la enfermedad, a análisis semanales, concluyendo que entre más información y más datos se obtengan para el preprocesamiento, mejor información se puede extraer.

Se detectó, de forma periódica, las zonas geográficas que presentaron una mayor incidencia de casos mediante el uso del software ArcGis y Google Maps permitiendo explorar, analizar y visualizar grandes conjuntos de datos de forma organizada, al disponer de una amplia variedad de herramientas visuales para su fácil interpretación, lo que ayudó a orientar campañas de medicina preventiva y acciones de contención de enfermedades de rápida propagación

Se desarrolló una metodología tanto objetiva como reproducible, lo que permitió proporcionar información a la Subdirección de Servicios Médicos.

El trabajo realizado implicó retos importantes como: considerando que, previo a esta investigación, se carecía de un método de detección de epidemias, **la validación de la metodología fue mediante los informes estadísticos enviados a la institución de salud.**

Un factor importante a destacar del uso de minería de datos es que debe tomarse como un apoyo para los especialistas, y no reemplazar el conocimiento que ellos, los resultados de los modelos no funcionan por sí solos, razón por la cual, los informes estadísticos fueron de suma importancia ya que era necesario que los expertos los interpretaran y validaran.

Se aplicó con éxito las técnicas de minería de datos para identificar los pacientes con características similares, esto se realizó por medio de Google Colaboratory, el cual facilitó en gran medida la preparación de los datos y, permitió realizar la aplicación de las técnicas de minería de datos de manera concisa para generar conocimiento útil a través de la caracterización de las enfermedades de cada uno.

Este último punto, abre muchas posibilidades, pues los pacientes se pueden tratar de manera distinta a través de una segmentación. Con la ayuda de la tecnología existente, se demostró que se pueden lograr mejoras en la calidad de este tipo de información.

5.2. Recomendaciones

Como parte de las recomendaciones que se sugieren a partir de esta investigación se encuentran:

- Utilizar distintas herramientas de software especializados en minería de datos y comparar tanto los resultados de un mismo algoritmo, como los de distintos algoritmos de minería de datos. Existe bibliografía que respalda que se puede llegar al mismo resultado por distintos medios.

- Debido a que la obtención del número de grupos en K-medias se realizó mediante pruebas con diferentes tamaños de núcleos, se recomienda aplicar indicadores con la validez necesaria para proporcionar el número óptimo de grupos formados.
- Se recomienda siempre realizar un muestreo preliminar de los datos, ya que esto definirá el modelo con el cual trabajar.
- En los registros epidemiológicos se presenta la información por enfermedad, donde viene registrado el número de paciente y la colonia en la cual habita. Es importante mencionar que el software ArcGIS cuando no reconoce una ubicación, la envía al centro del mapa, esto causa anomalías en los resultados, por lo que es preferible realizar una conversión y trabajar con coordenadas geográficas en lugar de los nombres específicos de las colonias.
- Otro factor importante, es que la institución de salud pública continúe elaborando periódicamente, en un intervalo sugerido de quince días, el reporte de enfermedades epidemiológicas descrito en la sección C del paso 3.3.1 planteado en la metodología, con el propósito de fundamentar y orientar acciones al seguir evaluando los distintos problemas de salud.
- Que las campañas de prevención de salud, por parte de las instituciones de salud, en las temporadas con mayor índice de casos, sean decisiones basadas en resultados, como árboles de decisión, además de enfocar sus esfuerzos en tratar de atender a aquellos pacientes con salud comprometida, como los encontrados por medio de k-means.

Las técnicas de minería de datos permiten extraer conocimiento que de modo manual no sería posible obtener. La exploración y análisis de datos para la detección de incidencia anormal de padecimientos en las diferentes zonas geográficas de la ciudad con ayuda de las herramientas de georreferenciación existentes nos revelan hacia dónde hay que encaminar los esfuerzos de prevención y atención, como en el caso de dengue y chikungunya, donde la Secretaría de Salud interviene periódicamente con fumigación y campañas para combatir y contener los brotes. La información que ha resultado de este

análisis es parte de la toma de decisiones para dirigir esfuerzos y recursos de modo que tengan un mayor impacto y efectividad.

5.3. Trabajos futuros

Con en este trabajo de tesis, se han comenzado a emplear técnicas de minería de datos para la detección de brotes de epidemias de rápida propagación en el municipio de Hermosillo, lo cual, simboliza un inicio para este tipo de investigación dentro del área de medicina preventiva en el municipio de Hermosillo, permitiendo diversas líneas de trabajo futuras.

En primera instancia, se recomienda utilizar distintas herramientas y tecnologías con el objetivo de mejorar los modelos de manera constante. Ningún modelo o algoritmo puede o debe ser usado de modo exclusivo, es necesario seguir explorando y comparando resultados en este tipo de investigación donde no existe el modelo o algoritmo “óptimo”, los mismos datos son los que dirigen la elección del o los modelos.

También es importante probar técnicas de clasificación y/o de asociación, con el propósito de no solo detectar sino predecir los brotes de epidemias. Este trabajo descriptivo solo fue el primer paso, es necesario mejorar constantemente el desarrollo, el bienestar social y la calidad de vida de la comunidad en general; una forma de lograrlo es con acceso a más datos, como análisis clínicos y medicamentos recetados, al igual que el registro de datos como peso, altura y presión arterial en cada consulta realizada por paciente.

6. REFERENCIAS

Akay, A., Dragomir, A. y Erlandsson, B.-E., 2015. Network-Based Modeling and Intelligent Data Mining of Social Media for Improving Care. *IEEE Journal of Biomedical and Health Informatics*, 19(1), pp. 210–218. doi: 10.1109/JBHI.2014.2336251.

Alias, H. *et al.*, 2014. Spatial distribution of malaria in Peninsular Malaysia from 2000 to 2009. *BioMed Central*. doi: 10.1186/1756-3305-7-186.

Becerra-Fernandez, I., Zanakis, S. H. y Walczak, S., 2002. Knowledge discovery techniques for predicting country investment risk, 43, pp. 787–800.

Berger, S., 2015. *Dengue: Global Status*. Los Angeles, California: GIDEON Informatics Inc.

Bonita, R., Beaglehole, R. y Kjellström, T., 2006. *Basic epidemiology*. 2da ed. *World Health Organization*. 2da ed. Ginebra: World Health Organization.

Brandt, K. G., Antunes, M. M. de C. y Pontes da Silva, G. A., 2015. Acute diarrhea : evidence-based management. *Jornal de Pediatria*. Sociedade Brasileira de Pediatria, 91(6), pp. S36–S43. doi: 10.1016/j.jpmed.2015.06.002.

Buczak, A. L. *et al.*, 2012. A data-driven epidemiological prediction method for dengue outbreaks using local and remote sensing data.

Buczak, A. L. *et al.*, 2014. Prediction of High Incidence of Dengue in the Philippines. *PLoS Neglected Tropical Diseases*, 8(4). doi: 10.1371/journal.pntd.0002771.

Buczak, A. L. *et al.*, 2015. Fuzzy association rule mining and classification for the prediction of malaria in South Korea. *BMC Medical Informatics and Decision Making*, 15(1), p. 47. doi: 10.1186/s12911-015-0170-6.

Cangrejo Aljure, D. y Agudelo, J. G., 2011. Minería de datos espaciales Spatial data mining – An overview. *Revista Avances en Sistemas e Informática*, 88(3), pp. 71–77. Disponible en: <http://www.redalyc.org/articulo.oa?id=133122679009>.

Carlson, A. A., Rose, T. N. y Gelinas, A., 2016. The Rundown : Management of Acute and Chronic Diarrhea. *DrugTopics Voice of the Pharmacist*, (June), pp. 55–64. Disponible en:

http://drugtopics.modernmedicine.com/sites/default/files/images/DrugTopics/UCONN/drt_p0616_CE.pdf.

Carroll, L. N. *et al.*, 2014. Visualization and analytics tools for infectious disease epidemiology: A systematic review. *Journal of Biomedical Informatics*. Elsevier Inc., 51, pp. 287–298. doi: 10.1016/j.jbi.2014.04.006.

Centers for Disease Control and Prevention, 2006. *Principles of epidemiology in public health practice: an introduction to applied epidemiology and biostatistics*. 3ra ed. Atlanta: U.S. Department of Health and Human Services. Disponible en: <https://stacks.cdc.gov/view/cdc/6914>.

Centers for Disease Control and Prevention, 2011. Preparedness and Response for Chikungunya Virus Introduction in the Americas. Washington, D.C: Pan Am J Public Health, p. 161. Disponible en: http://apps.who.int/iris/bitstream/10665/173322/1/Chikungunya_Virus.pdf.

Centers for Disease Control and Prevention, 2016. Diagnosis and Management of Tickborne Rickettsial Diseases: Rocky Mountain Spotted Fever and Other Spotted Fever Group Rickettsioses, Ehrlichioses, and Anaplasmosis — United States, 65(2).

Compieta, P. *et al.*, 2007. Exploratory spatio-temporal data mining and visualization. *Journal of Visual Languages and Computing*, 18(3), pp. 255–279. doi: 10.1016/j.jvlc.2007.02.006.

Cortés, F. M., Gómez, S. Y. y Ocazonez, R. E., 2007. Subtipos de virus dengue serotipos 2 , 3 y 4 aislados en el Departamento de Santander , Colombia, 59(3), pp. 186–192.

Crockett, D., Johnson, R. y Eliason, B., 2017. What is Data Mining in Healthcare? The Deployment System. *Health Catalyst*, pp. 1–13. Disponible en: <https://www.healthcatalyst.com/wp-content/uploads/2014/06/What-is-data-mining-in-healthcare.pdf>.

CropLife International, 2016. Vector Control. Brussels, Belgium: CropLife International. Disponible en: https://croplife.org/wp-content/uploads/pdf_files/Vector-Control-fact-sheet.pdf.

Dirección General de Epidemiología, 2016. Manual de procedimientos estandarizados para la vigilancia epidemiológica de las enfermedades transmitidas por vectores.

Dirección General de Promoción de la Salud, 2016. *Guía de Acciones Anticipatorias de Promoción de la Salud para las Enfermedades transmitidas por vector (Dengue, Chikungunya y Zika)*. Disponible en: <https://www.medbox.org/national-emergency-response-plans/guia-de-acciones-anticipatorias-de-promocion-de-la-salud-para-las-enfermedades-transmitidas-por-vector-dengue-chikungunya-y-zika/toolboxes/preview?>

Directorate General of Health Services, 2016. National guideline for clinical management of chikungunya. Delhi: National Vector Borne Disease Control Programme Directorate General of Health Services Ministry of Health & Family Welfare. Disponible en: <http://nvbdcp.gov.in/Doc/National-Guidelines-Clinical-Management-Chikungunya-2016.pdf>.

DPHHS, 2013. *The Common Cold. Health Guidelines*. Montana. Disponible en: <https://dphhs.mt.gov/Portals/85/dsd/documents/DDP/MedicalDirector/CommonCold101713.pdf>.

Dueñas, M. X., 2009. Minería de datos espaciales en búsqueda de la verdadera información. *SciELO Colombia*, pp. 137–156. Disponible en: <http://www.scielo.org.co/pdf/inun/v13n1/v13n1a07.pdf>.

Elebead, F. M. *et al.*, 2012. Mapping Cancer Disease Using Geographical Information System (GIS) in Gezira State-Sudan, pp. 830–839. doi: 10.1007/s10900-011-9517-9.

Fayyad, U., Piatetsky-shapiro, G. y Smyth, P., 1996. From Data Mining to Knowledge Discovery in Databases, 17(3), pp. 37–54.

Fayyad, U., Piatetsky-Shapiro, G. y Smyth, P., 1996. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11), pp. 27–34. doi: 10.1145/240455.240464.

Fernando Quevedo, R., 2011. Medidas de tendencia central y dispersión. *MEDwave*, (3), pp. 1–6. doi: 10.5867/medwave.2011.03.4934.

Gammino, V. M. *et al.*, 2014. Using Geographic Information Systems to Track Polio

Vaccination Team Performance: Pilot Project Report, 210(Suppl 1). doi: 10.1093/infdis/jit285.

Gibert, K., Sánchez-Marrè, M. y Codina, V., 2006. Elección de la técnica de minería de datos: Mapa conceptual de técnicas. *V Simposio de Teoría y Aplicaciones de Minería de Datos*, p. 7.

Gidey, G., Taju, S. y Hagos, A. S., 2006. *Introduction to Public Health*. 3ra ed. Mekelle University. Disponible en: http://www.dphu.org/uploads/attachements/books/books_2525_0.pdf.

Han, J., Kamber, M. y Pei, J., 2012. *Data Mining Concepts and Techniques*. 3ra ed. Waltham: Elsevier.

Hand, D., Mannila, H. y Padhraic, S., 2004. *Principles of Data Mining*. 2da ed. The Knowledge Engineering Review.

Hernández, J., Ramírez, M. J. y Ferri, C., 2004. *Introducción a la Minería de Datos*. Madrid: Pearson Prentice Hall.

Heymann, D. L., 2008. *Control of communicable diseases manual*. 19 ed. American Public Health Association. doi: 10.1016/S0022-3913(12)00047-9.

Huang, Y. C., 2013. The application of data mining to explore association rules between metabolic syndrome and lifestyles. *Health Information Management Journal*, 42(3), pp. 29–36. doi: 10.12826/18333575.2013.0008.

ISSSTESON, 2015. *ISSSTESON activa filtro contra Dengue y Chikungunya*. Disponible en: <http://www.isssteson.gob.mx/index.php/articulos/ver/> (Consultado: el 5 de enero de 2017).

Jensen, P. B., Jensen, L. J. y Brunak, S., 2012. Mining electronic health records: Towards better research applications and clinical care. *Nature Reviews Genetics*. Nature Publishing Group, 13(6), pp. 395–405. doi: 10.1038/nrg3208.

Kansas Department of Health and Environment, 2016. Tickborne Rickettsial Disease (TBRD) Investigation Guideline. *Kansas Disease Investigation Guidelines*. Kansas, pp. 1–

12. Disponible en:
http://www.kdheks.gov/epi/Investigation_Guidelines/Tickborne_Disease_Investigation_Guideline.pdf.

Khoury, M. J. *et al.*, 2013. Transforming Epidemiology for 21st Century Medicine and Public Health *Mun*, (1). doi: 10.1158/1055-9965.EPI-13-0146.

Kreis, I. A. *et al.*, 2013. *Essentials of Environmental Epidemiology for Health Protection*. Oxford University Press.

Kuri Morales, P., 2006. La vigilancia epidemiológica en México: elemento central para enfrentar problemas globales de salud.

Liu, Y. *et al.*, 2015. Spatio-temporal analysis of the relationship between climate and hand, foot, and mouth disease in Shandong province, China, 2008-2012. *BMC Infectious Diseases*, 15(1), pp. 1–8. doi: 10.1186/s12879-015-0901-4.

Loslier, L., 1994. Geographical Information Systems (GIS) from a Health Perspective, (September 1994), pp. 5–10.

Medecins Sans Frontieres, 2016. *Clinical Guidelines Diagnosis and treatment manual*. 2016 ed. Paris: Médecins Sans Frontières.

Miller, H. y Han, J., 2009. Geographic Data Mining and Knowledge Discovery: An Overview. en *Geographic Data Mining and Knowledge Discovery*. 2da ed. Boca Raton: CRC Press, pp. 1–26. doi: 10.1002/9780470690819.ch19.

Montgomery, D. C. y Runger, G. C., 2012. Probabilidad y estadística aplicadas a la ingeniería, p. 486. Disponible en:
<https://books.google.com/books?id=B2fkPQAACAAJ&pgis=1>.

Musa, G. J. *et al.*, 2013. Use of GIS Mapping as a Public Health Tool — From Cholera to Cancer, pp. 111–116. doi: 10.4137/HSI.S10471.TYPE.

National Center for Emerging and Zoonotic Infectious Diseases, 2012. Dengue and the *Aedes aegypti* mosquito. San Juan, Puerto Rico: Centers for Disease Control and Prevention, p. 2. Disponible en:

<https://www.cdc.gov/dengue/resources/30jan2012/aegyptifactsheet.pdf>.

Nobre, F. F. y Carvalho, M. S., 2016. Spatial and Temporal Analysis of Epidemiological Data, (September 1994), pp. 5–10.

Perumal, M. *et al.*, 2015. Spatial Data Mining Approaches for GIS – A Brief Review. *Advances in Intelligent Systems and Computing*, 338(April), pp. I–IV. doi: 10.1007/978-3-319-13731-5.

Pitrez, P. M. C. y Pitrez, J. L. B., 2003. Acute upper respiratory tract infections - outpatient diagnosis and treatment. *Jornal de Pediatria*, 79, pp. 77–86. Disponible en: <http://www.jped.com.br/conteudo/03-79-S77/ing.pdf>.

Riquelme, J. C., Ruiz, R. y Gilnert, K., 2006. Minería de Datos: Conceptos y Tendencias. *Inteligencia artificial: Revista Iberoamericana de Inteligencia Artificial*, 10(29), pp. 11–18. Disponible en: [https://idus.us.es/xmlui/bitstream/handle/11441/43290/Mineria de datos.pdf?sequence=1](https://idus.us.es/xmlui/bitstream/handle/11441/43290/Mineria_de_datos.pdf?sequence=1).

Rygielski, C., Wang, J. y Yen, D. C., 2002. Data mining techniques for customer relationship management. *Technology in society*, 24(4), pp. 483–502. Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.474.2713&rep=rep1&type=pdf>

Shafique, U. *et al.*, 2016. Data Mining in Healthcare for Heart Diseases. *International Journal of Innovation and Applied Studies*, 10(January), p. 1312.

Shafique, U. y Qaiser, H., 2014. A Comparative Study of Data Mining Process Models (KDD , CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, 12(1), pp. 217–222. Disponible en: <http://www.ijisr.issr-journals.org/>.

Shekhar, S. *et al.*, 2015. Spatiotemporal Data Mining: A Computational Perspective. *ISPRS International Journal of Geo-Information*, 4(4), pp. 2306–2338. doi: 10.3390/ijgi4042306.

Shickel, B. *et al.*, 2017. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. *IEEE Journal of Biomedical and Health Informatics*, pp. 1–16. doi: 10.1109/JBHI.2017.2767063.

Sirisena, P. *et al.*, 2017. Effect of Climatic Factors and Population Density on the Distribution of Dengue in Sri Lanka: A GIS Based Evaluation for Prediction of Outbreaks. *Plos One*, 12(1). doi: 10.1371/journal.pone.0166806.

Sistema Nacional de Vigilancia Epidemiológica, 2014. *Programa de Acción Específico*. Disponible en: https://www.gob.mx/cms/uploads/attachment/file/211946/PAE_2013-2018.pdf.

Sumathi, S. y Sivanandam, S. N., 2006. Introduction to Data Mining Principles. en *Introduction to Data Mining and its Applications*. Tamil Nadu: Springer-Verlag Berlin Heidelberg, pp. 1–20.

Weaver, S. C. y Lecuit, M., 2016. Chikungunya Virus and the Global Spread of a Mosquito-Borne Disease. *New England Journal of Medicine*, 372(13), pp. 1231–1239. doi: 10.1056/NEJMra1406035.

WHO, 2005. *The Treatment of Diarrhoea: A Manual for Physicians and Other Senior Health Workers*. 4ta ed. Ginebra: Department of Child and Adolescent Health and Development, World Health Organization. Disponible en: <http://apps.who.int/iris/bitstream/10665/43209/1/9241593180.pdf>.

Witten, I. H. y Frank, E., 2005. *Data Mining Practical Machine Learning Tools and Techniques*. 2da ed. San Francisco: Morgan Kaufmann. Disponible en: [ftp://ftp.ingv.it/pub/manuela.sbarra/Data Mining Practical Machine Learning Tools and Techniques - WEKA.pdf](ftp://ftp.ingv.it/pub/manuela.sbarra/Data%20Mining%20Practical%20Machine%20Learning%20Tools%20and%20Techniques%20-%20WEKA.pdf).

Yagoub, M. M., 2011. Geographie Information Systems (GIS) Application for Health : Case of Al Ain (UAE). *International Journal of Geoinformatics*, 7, pp. 21–28.

Zhou, H.-B. *et al.*, 2015. Spatio-temporal analysis of female breast cancer incidence in Shenzhen, 2007-2012. *Chinese journal of cancer*, 34(5), pp. 198–204. doi: 10.1186/s40880-015-0013-y.

7. ANEXOS

7.1. Anexo A

En este apartado se presenta la versión original de la plantilla llamada *Informe (Tema Intermedio)* del procesador de texto Word, de la cual se modificó únicamente la imagen.

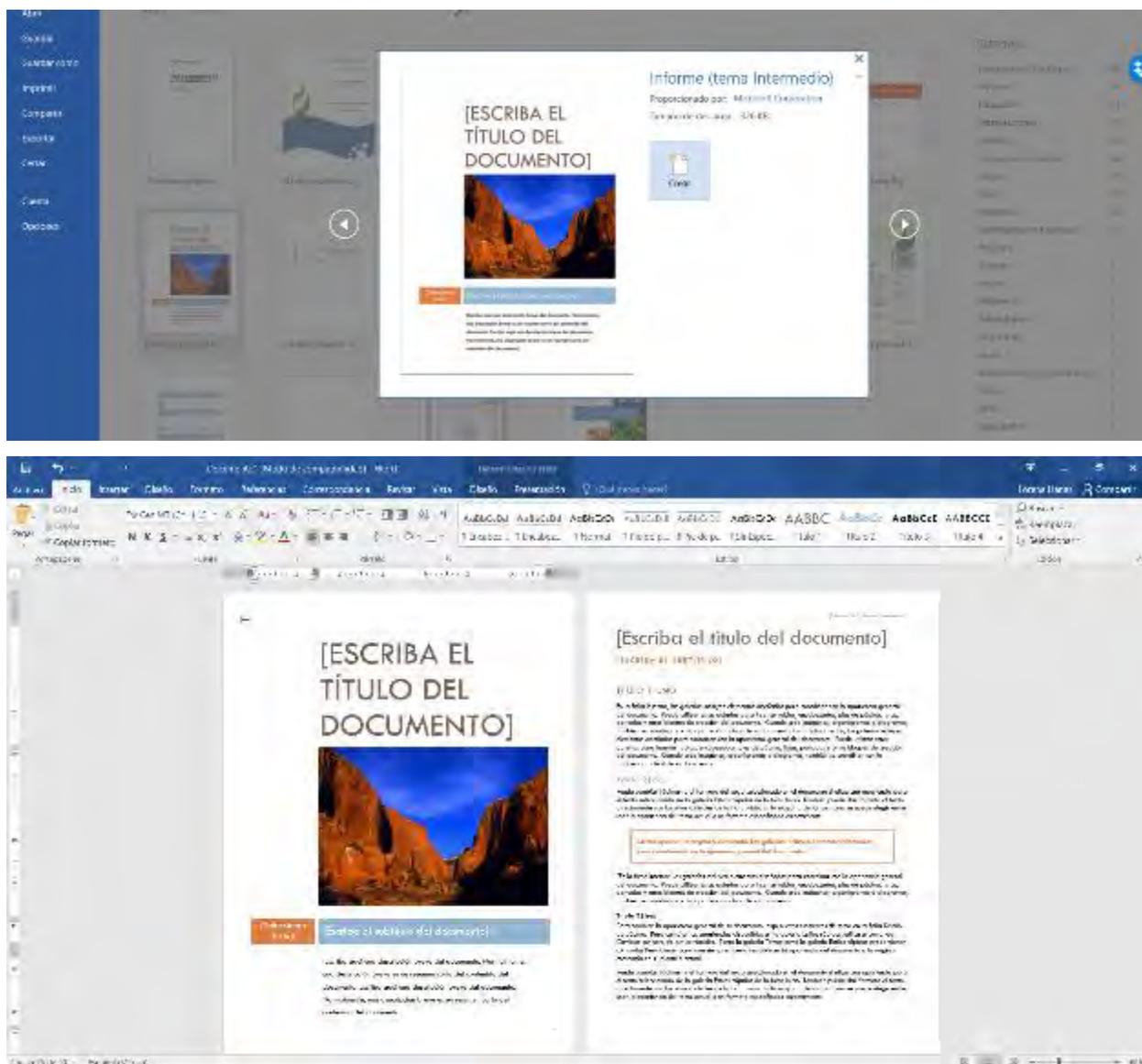
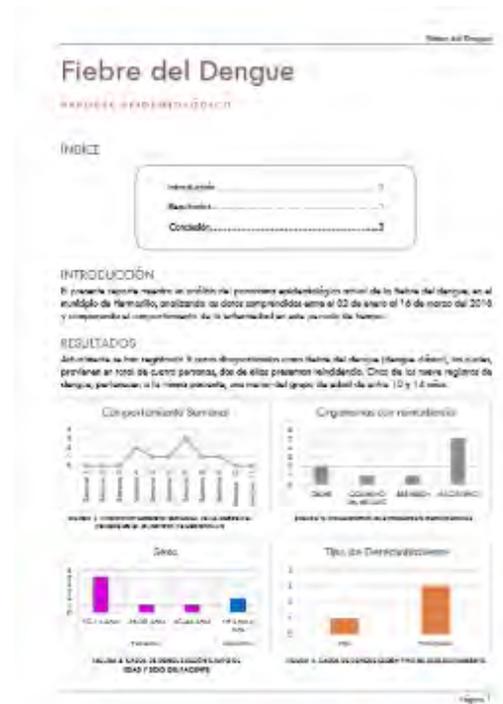


Figura 7.1. Plantilla Informe (Tema Intermedio) versión original

- A continuación, se muestra uno de los reportes enviados al Instituto de Salud Pública



7.2. Anexo B

En este anexo se enlistan las 753 colonias, que, al 20 de marzo del 2018, fecha en la cual se realizó la solicitud del documento, existen en el municipio de Hermosillo, siendo estas reconocidas por el H. Ayuntamiento de Hermosillo.



HERMOSILLO Tesorería

TESORERIA MUNICIPAL
DIRECCION DE CATASTRO
 "2018: AÑO DE LA SALUD"
 Hermosillo, Sonora a 20 de marzo del 2018.
 OF. TMH-DC-1200/03/18

C. LORENIA LLANES ROBLES
ING. EN MECATRÓNICA
UNIVERSIDAD DE SONORA
P R E S E N T E.

Con fundamento en el Artículo 129 de la Ley de Transparencia y Acceso a la Información Pública del Estado de Sonora, vengo en tiempo y forma legal a dar respuesta a su solicitud con número de folio 430918 hecha por correo electrónico donde pide lo siguiente:

" lista de las colonias que actualmente existen en el Municipio de Hermosillo "

Al respecto le informamos que existen 753 colonias:

22 DE SEPTIEMBRE	ALTA FIRENZE RESIDENCIAL
4 DE MARZO	ALTAMIRA
5 DE MAYO	ALTARES
ADOLFO DE LA HUERTA	ALTARES (SIN URBAN)
ADOLFO LOPEZ MATEOS	ALTARES II
ADOLFO LOPEZ MATEOS I ETAPA	ALTARIA RESIDENCIAL
ADOLFO LOPEZ MATEOS II ETAPA	ALTO VALLE
ADOLFO LOPEZ MATEOS III ETAPA	ALTOS DE ARAGON
ADOLFO LOPEZ MATEOS IV	ALVARO OBREGON
AGAVES RESIDENCIAL	AMPLIACION 4 DE MARZO
AGUALURCA	AMPLIACION EL MIRADOR
AGUAZARCA	AMPLIACION LADRILLERAS
AKIWIKI	AMPLIACION LAS PALMAS
ALBORADA	AMPLIACION PRIMERO HERMOSILLO
ALCALA RESIDENCIAL	AMPLIACION REAL DE MINAS
ALCATRAZ KINO VIEJO	ANGEL FLORES
ALEGRANZA RESIDENCIAL	ANTARA RESIDENCIAL
ALONDRAS RESIDENCIAL	APOLO
ALTA CALIFORNIA RESIDENCIAL	ARANDANOS RESIDENCIAL
ALTA CALIFORNIA RESIDENCIAL ETAPA IV	ARANJUEZ RESIDENCIAL



HERMOSILLO

Tesorería

**TESORERIA MUNICIPAL
DIRECCION DE CATASTRO**

"2018: AÑO DE LA SALUD"

Hermosillo, Sonora a 20 de marzo del 2018.

OF. TMH-DC-1200/03/18

ARBOLEDAS	CALIFORNIA II ETAPA
ARCADIA	CALIFORNIA III ETAPA
ARTESANOS KINO VIEJO	CALIFORNIA RESIDENCIAL
ATARDECERES	CAMINO DEL SERI
AZORES RESIDENCIAL	CAMINO REAL
BACHOCO	CAMINO REAL NORTE
BALDERRAMA	CAMPANARIO
BANUS	CAMPESTRE
BARCELONA RESIDENCIAL	CAMPESTRE LA HERRADURA
BARRIO DE LA CRUZ KINO VIEJO	CAMPESTRE LAS GRANJAS
BELLA VISTA	CAMPESTRE RESIDENCIAL
BELLAVISTA	CAMPO DE GOLF
BENEI RESIDENCIAL	CAMPOGRANDE RESIDENCIAL
BENITO JUAREZ	CANADA DE LOS NEGROS
BICENTENARIO RESIDENCIAL	CANTERAS RESIDENCIAL SECCION MONTALBAN
BILBAO RESIDENCIAL	CAPISTRANO
BONANZA RESIDENCIAL	CARDENO ETAPA I
BRETANA	CARMEN SERDAN
BUENA VISTA	CARNAVALITO KINO VIEJO
BUENA VISTA 2 SECCION AMANECER	CASA BLANCA
BUENA VISTA II ETAPA	CASA BONITA II
BUENAVENTURA	CASA BONITA RESIDENCIAL
BUENOS AIRES	CASA GRANDE RESIDENCIAL I
BUGAMBILIA	CASA GRANDE RESIDENCIAL I A Y B
CAFE COMBATE	CASA LINDA
ARBOLEDAS	CASA REAL
ARCADIA	CALIFORNIA II ETAPA
ARTESANOS KINO VIEJO	CALIFORNIA III ETAPA
ATARDECERES	CALIFORNIA RESIDENCIAL
AZORES RESIDENCIAL	CAMINO DEL SERI
BACHOCO	CAMINO REAL
BALDERRAMA	CAMINO REAL NORTE
BANUS	CAMPANARIO
CASALTA RESIDENCIAL	CATAVINA RESIDENCIAL
CASTELINA	CENTENARIO
CASTELLO RESIDENCIAL	CENTRAL DE ABASTOS
CATALINAS RESIDENCIAL	CENTRO



HERMOSILLO Tesorería

**TESORERIA MUNICIPAL
DIRECCION DE CATASTRO**

"2018: AÑO DE LA SALUD"
Hermosillo, Sonora a 20 de marzo del 2018.
OF. TMH-DC-1200/03/18

CENTRO KINO VIEJO	COUNTRY CLUB
CERESO	COUNTRY CLUB II
CERRADA CUCURPE	CRUZ GALVEZ
CERRADA DE MINAS	CUARTEL 15 A
CERRADA DIAMANTE	CUARTEL PITIC
CERRADA EL ENCANTO	CUARTEL XV
CERRADA LA CARIDAD	CUARTEL ZONA
CERRADA NORWALK	CUATRO OLIVOS
CERRADA PIMAS	CUAUHTEMOC
CERRO COLORADO	CUCURPE II
CERRO DE LA CAMPANA	CUMBRES RESIDENCIAL
CERRO DE LA CRUZ	CUMBRES RESIDENCIAL III
CHINO ABELARDO	DE LA UNION 17 DE JULIO
CHOYAL	DE LOS POETAS
CHULA VISTA	DIVISION DEL NORTE
CNOP	DUNAS
COLINA BLANCA	DUNAS II
COLINAS	EL APACHE
COLINAS DE SAN JAVIER	EL CONQUISTADOR RESIDENCIAL
COLINAS DEL MAR	EL CORTIJO
COLINAS II	EL CORTIJO PERIODISTA
COLOSO	EL CORTIJO UNISON
COMERCIAL CEDROS	EL DORADO I
COMPOSTELA RESIDENCIAL	EL DORADO II
CONDESA	EL ENCANTO
CONDominio HACIENDA REAL	EL ENCANTO II ETAPA
CONDominio SAN PABLO	EL ENCANTO RESIDENCIAL
CONDominio VILLAS DEL MAR	EL ESPLENDOR
CONDominios JACQUELYNN	EL ESPLENDOR III
CONJUNTO ARIZONA	EL JITO
CONQUISTADORES	EL LLANO
CONSTITUCION	EL MALECON
CORCELES RESIDENCIAL	EL MARIACHI
CORCELES RESIDENCIAL SECCION HIPICO	EL MEZQUITE
CORDOBA RESIDENCIAL	EL MIRADOR
CORONADO RESIDENCIAL	EL PINACATE II
CORSICA RESIDENCIAL	EL SAUCITO
COSTA DEL SOL	EL TAZAJAL


HERMOSILLO Tesorería

**TESORERIA MUNICIPAL
DIRECCION DE CATASTRO**

 "2018: AÑO DE LA SALUD"
 Hermosillo, Sonora a 20 de marzo del 2018.
 OF. TMH-DC-1200/03/18

EL TORREON	INALAMBRICA
EL TRIUNFO	INSURGENTES
EMILIANO ZAPATA	INTERNACIONAL
EQUIPAMIENTO EDUCATIVO	ISABELES
EUSEBIO KINO ISSSTESON	ISSSTE FEDERAL
FERROCARRILERA	ISSSTESON
FINCA LAS RIVERAS ETAPA I	ISSSTESON CENTENARIO
FIRENZE RESIDENCIAL	JACINTO LOPEZ
FLAMINGOS	JARDINES
FLORESTA SECC VILLA ALMENDRAS	JARDINES DE MONACO
FLORESTA SECC VILLA CEIBA	JEREZ DEL VALLE
FLORESTA SECC VILLA CIPRESES	JESUS GARCIA
FLORESTA SECC VILLA CIRUELOS	JESUS GARCIA I
FLORESTA SECC VILLA FRESNOS	JESUS GARCIA II
FLORESTA SECC VILLA SAUCES	JORGE VALDEZ MUNOZ
FONHAPO	JOYAS DEL BACHOCO
FOVISSSTE I Y II	KINO BAY ESTATES
FRACCIONAMIENTO SENDA DORADA	LA ANTORCHA
FRANCISCO EUSEBIO KINO	LA CAMPANA
FRANCISCO VILLA	LA CAMPIDA
FUENTE DE PIEDRA	LA CAMPIDA III IV Y V ETAPAS
FUENTES DEL CENTENARIO	LA CARIDAD
FUENTES DEL MEZQUITAL	LA CHOLLA
GALA	LA CHOLLA II SECCION
GALA II	LA CIMA
GALA III	LA CIMA II
GENOVA RESIDENCIAL	LA CORUÑA
H AYUNTAMIENTO	LA ENCANTADA PRIVADA ALEBRIJES
HACIENDA DE LA FLOR	LA ENCANTADA PRIVADA AZULEJOS
HACIENDA DEL SOL	LA ENCANTADA PRIVADA LEYENDAS
HACIENDA LOS ALCATRACES	LA ENCANTADA PRIVADA TESOROS
HACIENDA LOS LIRIOS	LA HUERTA
HACIENDA RESIDENCIAL CONDOMINAL	LA JOLLA VILLA DE LAS ESMERALDAS
HACIENDAS DEL SUR	LA JOLLA VILLA DE LAS TURQUESAS
HEBERTO CASTILLO	LA JOLLA VILLA DE LAS TURQUESAS ETAPA 2
HECTOR FCO ROMERO NUEVA VICTORIA	LA JOLLA VILLA DE LAS TURQUESAS ETAPA 3
HECTOR FRANCISCO ROMERO ETAPA II	LA JOLLA VILLA DE LOS DIAMANTES
IBIZA RESIDENCIAL	LA JOLLA VILLA DE LOS DIAMANTES ETAPA 2



HERMOSILLO

Tesorería

**TESORERIA MUNICIPAL
DIRECCION DE CATASTRO**

"2018: AÑO DE LA SALUD"

Hermosillo, Sonora a 20 de marzo del 2018.
OF. TMH-DC-1200/03/18

LA JOLLA VILLA DE LOS ZAFIROS	LAS LOMAS PRIVADA DEL BOSQUE
LA MANGA	LAS LOMAS PRIVADA LAS LOMAS
LA MATANZA	LAS LOMAS REAL DEL PARQUE
LA METALERA	LAS LOMAS SECC BONITA
LA MOSCA	LAS LOMAS SECC CASTANOS
LA PALOMA RESIDENCIAL I	LAS LOMAS SECC LOS ALMENDROS
LA PALOMA RESIDENCIAL II	LAS LOMAS SECC LOS MANZANOS
LA PLANCHA	LAS LOMAS SECC SEC
LA PONDEROSA	LAS LOMAS SECCION EL PARAISO
LA PRIMAVERA	LAS MINITAS
LA RIOJA	LAS PALMAS
LA RIOJA NORTE	LAS PEREDAS
LA SARDINA	LAS PILAS
LA VENTANA (ARCO IRIS)	LAS PLAZAS
LA VERBENA	LAS PRADERAS IV
LA VERBENA SECCION ASTURIAS I ETAPA	LAS PRADERAS NORTE
LA VERBENA SECCION ASTURIAS II ETAPA	LAS PRADERAS SUR (FOVISSSTE)
LA VICTORIA	LAS PROVINCIAS
LADRILLERAS	LAS QUINTAS
LAGO ESCONDIDO	LAS RIBERAS DEL PITIC SECCION VILLA DEL PARQUE
LAS ACACIAS	LAS TORRES
LAS ALAMEDAS	LAS VEREDAS
LAS AMAPOLAS NORTE	LAS VILLAS
LAS AMAPOLAS SUR	LAS VILLAS DEL NOGAL
LAS ARBOLEDAS	LAS VISTAS
LAS AVES	LAURA ALICIA FRIAS DE LOPEZ NOGALES
LAS CARRETAS	LEY 57
LAS CUEVAS	LIBERTAD
LAS DILIGENCIAS	LIRIOS
LAS FUENTES	LLANO VERDE
LAS GLORIAS	LOMA LINDA
LAS GRANJAS	LOMA LINDA KINO VIEJO
LAS HACIENDAS	LOMAS ALTAS
LAS LOMAS	LOMAS DE BARROCAN
LAS LOMAS DE CORTES	LOMAS DE KINO
LAS LOMAS DEL MIRADOR	LOMAS DE LINDA VISTA
LAS LOMAS DEL SOL	LOMAS DE MADRID
LAS LOMAS DEL SUR	LOMAS DE REFORMA



HERMOSILLO

Tesorería

**TESORERIA MUNICIPAL
DIRECCION DE CATASTRO**

"2018: AÑO DE LA SALUD"

Hermosillo, Sonora a 20 de marzo del 2018.

OF. TMH-DC-1200/03/18

LOMAS DE SAN GERMAN	MESAVERDE
LOMAS DEL NORTE Y PARAISO	METROCENTRO
LOMAS DEL PITIC	MEZQUITAL DEL ORO
LOMAS PITIC	MIGUEL ANGEL JIMENEZ
LOPEZ PORTILLO	MIGUEL HIDALGO
LORETO	MINIPARQUE INDUSTRIAL
LOS ALAMOS	MIRASOLES
LOS ALTOS RESIDENCIAL	MISION
LOS ANGELES	MISION DE SAN IGNACIO
LOS ANGELES X ETAPA	MISION DEL ARCO
LOS ARBOLITOS	MISION DEL REAL
LOS ARCOS	MISION DEL SOL
LOS ARROYOS	MISIONEROS
LOS CANARIOS	MODELO
LOS GIRASOLES	MONACO PRIVADAS RESIDENCIALES
LOS GRANADOS DEL LLANO	MONTECARLO
LOS LAGOS	MONTEREAL
LOS MANANTIALES	MONTEROSA ETAPA III
LOS NARANJOS	MONTEROSA ETAPA III B
LOS OLIVOS	MONTEROSA RESIDENCIAL
LOS PINOS RESIDENCIAL	MONTEROSA SEGUNDA ETAPA
LOS PORTALES	MURANO
LOS ROSALES	NACAMERI
LOS SABINOS	NATURA
LOS SAUCES	NAZAREO RESIDENCIAL
LOS VIDEDOS	NORBERTO ORTEGA
LUIS DONALDO COLOSIO 3RA ETAPA	NUEVA CASTILLA
LUIS DONALDO COLOSIO 4TA ETAPA	NUEVA ESPANA
LUIS ENCINAS	NUEVA PALMIRA
MALLORCA RESIDENCIAL	NUEVA PALMIRA PRIVADA PALMIRA
MANUEL GOMEZ MORIN	NUEVA VICTORIA 2
MANUEL J CLOUTHIER	NUEVO DESIERTO
MARGARITAS OLIVOS	NUEVO HERMOSILLO
MARIN PARAISO	NUEVO HORIZONTE
MARSELLA RESIDENCIAL	NUEVO PROGRESO
MARTIRES DE CANANEA	NUEVO SAHUARO
MEDITERRANEO	OASIS DE KINO
MEDITERRANEO RESIDENCIAL	OASIS DEL SOL



TESORERIA MUNICIPAL
DIRECCION DE CATASTRO
 "2018: AÑO DE LA SALUD"
 Hermosillo, Sonora a 20 de marzo del 2018.
 OF. TMH-DC-1200/03/18

OASIS LANTANA	PASEO SAN ANGEL II SECC CERRADA LAS PLACITAS
OASIS PALMEIRA	PASEO SAN ANGEL II SECC LAS PLACITAS I
OASIS PALMEIRA II	PASEO SAN ANGEL II SECC LAS PLACITAS II
OASIS SANTA FE	PASEO SAN ANGEL II SECC LAS PLACITAS V
OASIS SOLANA	PASEO SAN ANGEL III SECC
OBISPOS II	PASEOS DEL PEDREGAL
OBISPOS RESIDENCIAL	PASEOS DEL PEDREGAL II
OFICINAS FEDERALES AL OESTE	PEDREGAL DE LA VILLA
OLIVARES	PEMEX
OPATAS	PERIODISTA
PALERMO SECCION FERRARA	PERISUR
PALERMO SECCION PARTANNA	PILARES
PALERMO SECCION VENETO	PIMENTEL
PALMA DORADA RESIDENCIAL	PITIC
PALMAR DEL SOL	PLAZA DE LAS FRANQUICIAS
PALO FIERRO	PLAZA GRANDE
PALO VERDE	PLAZA REAL
PALO VERDE INDEUR	POPULAR
PARADA CANO	PORTAL DE ROMANZA
PARQUE CENTRAL	PORTAL DEL LAGO
PARQUE INDUSTRIAL	PORTAL DEL PITIC
PARQUE INDUSTRIAL	PORTON DEL ARCO
PARQUE INDUSTRIAL DE HERMOSILLO	POTRERO DE BUSHARD
PARQUE INDUSTRIAL DE SONORA	PRADOS DE BUGAMBILIAS
PARQUE INDUSTRIAL DYNATECH	PRADOS DE BUGAMBILIAS II
PARQUE INDUSTRIAL DYNATECH SUR	PRADOS DEL CENTENARIO
PARQUE INDUSTRIAL HERMOSILLO NORTE	PRADOS DEL SOL
PARQUE INDUSTRIAL LA LABOR	PRIMAVERA KINO VIEJO
PARQUE INDUSTRIAL LATITUD	PRIMERO HERMOSILLO
PARQUE INDUSTRIAL PIMEX	PRIVADA CASA BLANCA
PASEO DE LA COLINA	PRIVADA CORDOVA
PASEO DE LAS MISIONES	PRIVADA PALMIRA RESIDENCIAL ETAPA II
PASEO DE LAS PALMAS	PRIVADA QUINTA EMILIA
PASEO DEL PALMAR	PRIVADA RESIDENCIAL LA CAÑADA
PASEO DEL SOL	PRIVADA ROSALES
PASEO MARGARITAS	PRIVADA SAN GERMAN
PASEO REAL RESIDENCIAL	PRIVADA VICTORIA
PASEO SAN ANGEL	PRIVADAS DEL BOSQUE



HERMOSILLO Tesorería

**TESORERIA MUNICIPAL
DIRECCION DE CATASTRO**

"2018: AÑO DE LA SALUD"
Hermosillo, Sonora a 20 de marzo del 2018.
OF. TMH-DC-1200/03/18

PRIVADAS DEL CID RESIDENCIAL	PUERTA REAL RESIDENCIAL IX
PRIVADAS DEL REAL	PUERTA REAL RESIDENCIAL VII ETAPA
PRIVADAS DEL REAL II SECC	PUERTA REAL RESIDENCIAL VII SEGUNDA SECCION
PRIVADAS DEL REAL III SECC	PUERTA REAL RESIDENCIAL VIII
PRIVADAS DEL REAL SECC NORTE	PUERTA REAL V
PRIVADAS DEL REY	PUERTA REAL VI
PRIVADAS DEL SAUCE	PUERTA REAL VI 2da SECCION
PRIVADAS RESIDENCIALES	PUERTA REAL X
PROGRESISTA	PUESTA DEL SOL
PROGRESO KINO VIEJO	PUNTA BELLA
PROYECTO RIO SONORA	PUNTA DE LAGO
PUEBLITOS CDA GUARIJIOS	QUINTA EMILIA
PUEBLITOS CDA KIKAPU	QUINTA ESMERALDA ETAPAS 1 Y 2
PUEBLITOS CDA KUKAPU	QUINTA LA PALOMA
PUEBLITOS CDA PAPAGOS	QUINTA REAL
PUEBLITOS SECC LOS ALAMOS	QUINTAS DEL SOL ETAPA 10
PUEBLO ALEGRE	QUINTAS DEL SOL ETAPAS II V VII X
PUEBLO ALTO	QUINTAS DEL SOL IV ETAPA
PUEBLO BONITO	QUINTAS DEL SOL RESIDENCIAL I ETAPA
PUEBLO DEL ANGEL	QUINTAS DEL SOL RESIDENCIAL III ETAPA
PUEBLO DEL ORO	QUINTAS DEL SOL RESIDENCIAL VIII ETAPA
PUEBLO DEL SOL	QUINTAS DEL SOL VI ETAPA
PUEBLO DORADO	QUINTAS DEL SOL XI Y XII
PUEBLO ESCONDIDO	QUINTAS GALICIA
PUERTA DE HIERRO	RACQUET CLUB I SECC SUR
PUERTA DE HIERRO II	RACQUET CLUB II SECC NORTE
PUERTA DEL REY	RANCHETES SANTA LUCIA
PUERTA DEL REY ETAPA II	RANCHITO
PUERTA DEL REY ETAPA III	RANCHO BONITO
PUERTA DEL REY ETAPA IV	RANCHO SAN IGNACIO
PUERTA DEL REY ETAPA V	REAL DE MINAS
PUERTA DEL REY VI	REAL DE MONTEJO
PUERTA DEL SOL	REAL DE QUIROGA
PUERTA ESMERALDA	REAL DE SEVILLA
PUERTA GRANDE	REAL DE SEVILLA II ETAPA
PUERTA REAL RESIDENCIAL	REAL DE SEVILLA III ETAPA 1ERA SECCION
PUERTA REAL RESIDENCIAL III	REAL DE SEVILLA III ETAPA 2DA SECCION
PUERTA REAL RESIDENCIAL IV	REAL DE TOLEDO



HERMOSILLO

Tesorería

**TESORERIA MUNICIPAL
DIRECCION DE CATASTRO**

 "2018: AÑO DE LA SALUD"
 Hermosillo, Sonora a 20 de marzo del 2018.
 OF. TMH-DC-1200/03/18

REAL DEL ALAMITO PRIMERA SECCION	SALVADOR ALVARADO
REAL DEL ALAMITO SECCION RIO BONITO	SALVATIERRA RESIDENCIAL
REAL DEL ALAMITO SECCION Z	SAN ANDRES RESIDENCIAL
REAL DEL BOSQUE	SAN ANGEL
REAL DEL CARDO	SAN ANTONIO
REAL DEL CARMEN II	SAN BENITO
REAL DEL CARMEN III	SAN BERNARDINO ETAPA 1
REAL DEL CARMEN IV	SAN BERNARDINO ETAPA 2
REAL DEL CARMEN RESIDENCIAL	SAN BERNARDINO ETAPA 3
REAL DEL CARMEN V SECCION	SAN BERNARDINO ETAPA 4
REAL DEL COBRE	SAN BOSCO
REAL DEL LLANO	SAN DIEGO
RENACIMIENTO	SAN FRANCISCO
RESIDENCIAL DE ANZA	SAN ISIDRO
RESIDENCIAL KINO	SAN JAVIER
RESIDENCIAL LOS SANTOS	SAN JERONIMO
RESIDENCIAL MONTEBELLO	SAN JUAN
RESIDENCIAL NAVARRETE	SAN JUAN BOSCO II ETAPA
RESIDENCIAL PARQUE VERSALLES	SAN JUDAS TADEO
RESIDENCIAL PENASCO	SAN LORENZO
RESIDENCIAL PIMA	SAN LUIS
RESIDENCIAL RIO SECO	SAN LUIS REY
RETORNO DE LA MISION	SAN MARCOS
REVOLUCION	SAN MARCOS II
REVOLUCION II	SAN PABLO
RINCON AGRADABLE	SAN PEDRO
RINCON DE LA HERRADURA	SAN SEBASTIAN SECC VILLA BONITA
RINCON DE PALMAS	SANTA ANITA RESIDENCIAL
RINCONADA DE LA CRUZ	SANTA BARBARA
RIO GRANDE	SANTA BARBARA SECCION DEL MIRADOR
RIVELLO RESIDENCIAL	SANTA CRUZ RESIDENCIAL
ROMANZA RESIDENCIAL	SANTA FE
ROSARIO IBARRA DE PIEDRA	SANTA ISABEL
SACRAMENTO	SANTA LUCIA
SAHUARO	SANTA MARIA
SAHUARO FINAL	SANTA MONICA
	SANTERRA RESIDENCIAL
SAHUARO INDECO	SECTOR A


HERMOSILLO Tesorería

TESORERIA MUNICIPAL
DIRECCION DE CATASTRO

"2018: AÑO DE LA SALUD"

Hermosillo, Sonora a 20 de marzo del 2018.

OF. TMH-DC-1200/03/18

SECTOR B	TUBAC
SECTOR C	TULIPANES
SECTOR D	TUMACACORI
SEMINARIO RESIDENCIAL	UNION DE COLONOS
SIENA RESIDENCIAL	UNION DE LADRILLEROS
SIERRA BONITA	UNION DE PRODUCTORES DE LADRILLO CERRO COLORADO PONIENTE A C
SIERRA CLARA	UNIVERSIDAD
SIERRA VISTA	UNIVERSIDAD DE SONORA
SOLEI ETAPA IV Y V	URBI ALAMEDA LOS ENCINOS RESIDENCIAL II
SOLEIL RESIDENCIAL	URBI ALAMEDA LOS FRESNOS
SOLERA	URBI VILLA CAMPESTRE
SOLIDARIDAD	URBI VILLA CAMPESTRE III
SOLIDARIDAD IV ETAPA	URBI VILLA DEL CEDRO SECCION 2
SOLIDARIDAD OCCIDENTAL	URBI VILLA DEL PRADO
SOLIDARIDAD SEXTA ETAPA	URBI VILLA DEL REY II SECCION ETAPA I
SOLIDARIDAD V ETAPA	URBI VILLA DEL REY II SECCION ETAPA II
SOLIDARIDAD V FRANCISCO ROMERO	URBI VILLA DEL REY II SECCION ETAPA III
SOLIDARIDAD VII ETAPA	URBI VILLA DEL REY III SECCION
SOLIDARIDAD VIII ETAPA	URBI VILLA DEL REY SECCION CASTANOS II
SOLIDARIDAD VILDOSOLA (ZONA SUR)	URBI VILLA DEL REY SECCION CASTANOS III
SONACER I	URBI VILLA DEL REY SECCION CASTANOS IV
SONACER II	URBI VILLA DEL REY SECCION LOS ALMENDROS II
SONACER SECCION BARCELO	URBI VILLA DEL REY SECCION LOS ALMENDROS III
SONOMEX NORTE	URBI VILLA DEL REY SECCION LOS MANZANOS II
STANZA SOLARE RESIDENCIAL	VALLE BONITO
STAU UNIVERSIDAD	VALLE DE ARANDAS RESIDENCIAL
TABACHINES	VALLE DE LOS ALMENDROS
TERRABLANCA	VALLE DE SANTA MONICA
TERRANOVA	VALLE DEL LAGO
TERRAZAS DEL MIRADOR	VALLE DEL LAGO II
TIERRA COLORADA	VALLE DEL LAGO V
TIERRA NUEVA	VALLE DEL MARQUEZ
TIRO AL BLANCO	VALLE DEL PORTAL
TIROCAPES	VALLE ESCONDIDO
TOPACIO RESIDENCIAL	VALLE GRANDE
TORRE DE PIEDRA	VALLE HERMOSO
TORREPLATA	VALLE SANTA MONICA
TOSALI RESIDENCIAL	VALLE VERDE



TESORERIA MUNICIPAL
DIRECCION DE CATASTRO
 "2018: AÑO DE LA SALUD"
 Hermosillo, Sonora a 20 de marzo del 2018.
 OF. TMH-DC-1200/03/18

VELAS RESIDENCIAL	VILLAS DEL PALMAR
VILLA BONITA	VILLAS DEL REY
VILLA CALIFORNIA	VILLAS DEL SUR
VILLA COLONIAL	VILLAS RESIDENCIAL
	VINEDOS
	VIRREYES
VILLA DE LOS CORCELES	VISTA DEL LAGO
VILLA DE PARRAS	VISTA DEL SOL RESIDENCIAL
VILLA DE SERIS	VISTA REAL
VILLA DEL PITIC	Y GRIEGA
VILLA DEL REAL	ZONA INDUSTRIAL
VILLA DORADA	
VILLA FLORENCIA	
VILLA FONTANA	
VILLA GUADALUPE	
VILLA HERMOSA	
VILLA LOURDES	
VILLA MERCEDES	
VILLA MERLOT RESIDENCIAL	
VILLA RESIDENCIAL BONITA IX X XI Y XII	
VILLA SATELITE	
VILLA SOL	
VILLA SONORA	
VILLA SURENA	
VILLA TOSCANA RESIDENCIAL	
VILLA VERDE	
VILLA VERDE ETAPA II	
VILLA VERDE III	
VILLA VERDE IV	
VILLA VERDE V	
VILLAS DE CORTES	
VILLAS DE ROMANZA	
VILLAS DE TURIN	
VILLAS DEL CORTIJO	
VILLAS DEL MEDITERRANEO	
VILLAS DEL MEDITERRANEO II	
VELAS RESIDENCIAL	
VILLAS DEL MEDITERRANEO III	



TESORERIA MUNICIPAL
DIRECCION DE CATASTRO
"2018: AÑO DE LA SALUD"
Hermosillo, Sonora a 20 de marzo del 2018.
OF. TMH-DC-1200/03/18

Esperando satisfacer su derecho de acceso a la Información Pública y sin otro particular quedo a sus órdenes y reciba de antemano un cordial saludo.

ATENTAMENTE


ING. DAVID RAMOS FELIX
Director de Catastro
Tesorería Municipal del Ayuntamiento de Hermosillo



AYUNTAMIENTO DE HERMOSILLO
TESORERIA MUNICIPAL
DIRECCION DE CATASTRO

C.c.p. archivo

7.3. Anexo C

- Agrupamiento con K= 2

```
cluster 0 Registros: 24783
=====
d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.047451882338699916
d18 DIABETES MELLITUS NO INSULINODEPENDIENTE,SIN MENCION DE COMPLICACION : 0.1854900536658193
d28 FARINGITIS AGUDA : 1.104910624218214

cluster 1 Registros: 4170
=====
d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.004796163069544364
d4 DIABETES MELLITUS ASOCIADA CON DESNUTRICION,CON COMPLICACIONES CIRCULATORIAS PERIFERICAS : 0.010071942446043165
d21 DIARREA CON DESHIDRATAACION : 0.06258992805755395
d28 FARINGITIS AGUDA : 2.551318944844125
d56 RINOFARINGITIS AGUDA [RESFRIADO COMUN] : 4.509592326139089
```

- Agrupamiento con K=5

```
cluster 0 Registros: 7018
=====
d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.04887432316899402
d8 DIABETES MELLITUS INSULINODEPENDIENTE (DM) : 2.445141065830721

cluster 1 Registros: 8255
=====
d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.04227740763173834
d8 DIABETES MELLITUS INSULINODEPENDIENTE (DM) : 0.5189582071471835
d36 FARINGOAMIGDALISTIS : 2.632344033918837

cluster 2 Registros: 1409
=====
d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.10078069552874379
d8 DIABETES MELLITUS INSULINODEPENDIENTE (DM) : 5.853087295954578

cluster 3 Registros: 1276
=====
d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.04623824451410658
d8 DIABETES MELLITUS INSULINODEPENDIENTE (DM) : 1.9647335423197492
d36 FARINGOAMIGDALISTIS : 6.801724137931035

cluster 4 Registros: 9596
=====
d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.07721967486452688
d8 DIABETES MELLITUS INSULINODEPENDIENTE (DM) : 0.5911838265944144
d36 FARINGOAMIGDALISTIS : 0.6360983743226344
```

- Agrupamiento con K=10

```

cluster 0 Registros: 1485
=====
d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.0026936026936026937
d4 DIABETES MELLITUS ASOCIADA CON DESNUTRICION,CON COMPLICACIONES CIRCULATORIAS PERIFERICAS : 0.00404040404040404
d18 DIABETES MELLITUS NO INSULINODEPENDIENTE,SIN MENCION DE COMPLICACION : 0.006060606060606061
d21 DIARREA CON DESHIDRATAACION : 0.0430976430976431
d28 FARINGITIS AGUDA : 1.606060606060606
d56 RINOFARINGITIS AGUDA [RESFRIADO COMUN] : 6.5111111111111111

cluster 1 Registros: 8498
=====
d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.006001412096963992
d18 DIABETES MELLITUS NO INSULINODEPENDIENTE,SIN MENCION DE COMPLICACION : 0.01070840197693575
d21 DIARREA CON DESHIDRATAACION : 0.04306895740174159
d28 FARINGITIS AGUDA : 0.6216756883972699
d56 RINOFARINGITIS AGUDA [RESFRIADO COMUN] : 2.4976465050600143

cluster 2 Registros: 5403
=====
d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.00518230612622617
d3 DIABETES MELLITUS ASOCIADA CON DESNUTRICION,CON COMA : 0.005367388487877105
d4 DIABETES MELLITUS ASOCIADA CON DESNUTRICION,CON COMPLICACIONES CIRCULATORIAS PERIFERICAS : 0.0229502128447159
d21 DIARREA CON DESHIDRATAACION : 0.0459004256894318
d28 FARINGITIS AGUDA : 2.3751619470664447

cluster 3 Registros: 1900
=====
d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.0005263157894736842
d2 DIABETES MELLITUS ASOCIADA CON DESNUTRICION,CON CETOACIDOSIS : 0.0015789473684210526
d3 DIABETES MELLITUS ASOCIADA CON DESNUTRICION,CON COMA : 0.0036842105263157894
d4 DIABETES MELLITUS ASOCIADA CON DESNUTRICION,CON COMPLICACIONES CIRCULATORIAS PERIFERICAS : 0.015789473684210527
d18 DIABETES MELLITUS NO INSULINODEPENDIENTE,SIN MENCION DE COMPLICACION : 0.016842105263157894
d21 DIARREA CON DESHIDRATAACION : 0.07526315789473684
d28 FARINGITIS AGUDA : 4.142631578947368

cluster 4 Registros: 45
=====
d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.044444444444444446
d8 DIABETES MELLITUS INSULINODEPENDIENTE (DM) : 0.11111111111111111
d15 DIABETES MELLITUS NO INSULINODEPENDIENTE,CON COMPLICACIONES NO ESPECIFICADAS : 0.13333333333333333
d17 DIABETES MELLITUS NO INSULINODEPENDIENTE,CON OTRAS COMPLICACIONES ESPECIFICADAS : 1.0
d18 DIABETES MELLITUS NO INSULINODEPENDIENTE,SIN MENCION DE COMPLICACION : 22.844444444444445

cluster 5 Registros: 36
=====
d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 16.555555555555557

cluster 6 Registros: 1308
=====
d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.00382262996941896
d4 DIABETES MELLITUS ASOCIADA CON DESNUTRICION,CON COMPLICACIONES CIRCULATORIAS PERIFERICAS : 0.017584097859327217
d18 DIABETES MELLITUS NO INSULINODEPENDIENTE,SIN MENCION DE COMPLICACION : 0.04434250764525994
d28 FARINGITIS AGUDA : 0.5344036697247706
d40 FIEBRE DEL DENGUE [DENGUE CLASICO] : 3.0863914373088686

cluster 7 Registros: 280
=====
d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.08928571428571429
d8 DIABETES MELLITUS INSULINODEPENDIENTE (DM) : 0.09285714285714286
d10 DIABETES MELLITUS INSULINODEPENDIENTE, SIN MENCION DE COMPLICACION : 0.21428571428571427
d17 DIABETES MELLITUS NO INSULINODEPENDIENTE,CON OTRAS COMPLICACIONES ESPECIFICADAS : 0.975
d18 DIABETES MELLITUS NO INSULINODEPENDIENTE,SIN MENCION DE COMPLICACION : 6.8

```

cluster 8 Registros: 9590

=====

d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.0502606882168926
d18 DIABETES MELLITUS NO INSULINODEPENDIENTE,SIN MENCION DE COMPLICACION : 0.14327424400417102
d28 FARINGITIS AGUDA : 0.5686131386861314
d56 RINOFARINGITIS AGUDA [RESFRIADO COMUN] : 0.608133472367049

cluster 9 Registros: 408

=====

d1 DIABETES MELLITUS ASOCIADA CON DESNUTRICION : 0.004901960784313725
d18 DIABETES MELLITUS NO INSULINODEPENDIENTE,SIN MENCION DE COMPLICACION : 0.0196078431372549
d21 DIARREA CON DESHIDRATAACION : 0.12990196078431374
d28 FARINGITIS AGUDA : 8.318627450980392