UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

Facultad 3

Método no supervisado para la selección de rasgos en problemas de regionalización.

Tesis presentada para optar por el título de Ingeniero en Ciencias Informáticas

Autor: Monica Frómeta Torres

Tutor: Msc. Liset González Polanco

Yadian Guillermo Pérez Betancourt

La Habana, agosto del 2020

DECLARACIÓN DE AUTORÍA

Declaro ser autor de la presente tesis y se reconoce a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo. Para que así conste se firma la presente a los ____ días del mes de agosto del año 2020.

Monica Frómeta Torres

_____

Msc. Liset González Polanco

_____

Msc. Yadian Guillermo Pérez Betancourt

_____

AGRADECIMIENTOS

A mis padres, Abdias e Idalmis por formarme, enseñarme y amarme desde mis primeros momentos.

A Yadira, gracias por todo tu amor y tus consejos.

A mi hermana, gracias por estar ahí para mí, por tus risas y ocurrencias.

A mis abuelos, Mirta, Enrique, Patricio y Elsa; fuente de sabiduría y amor incondicional.

A mi Martín, gracias mi amor por tu compañía, tu amor y tu apoyo en este y muchos otros proyectos.

A mis tutores, Yadian y Lisset, gracias por brindarme su guía, su experiencia y toda su ayuda.

A mi pandilla, Amanda, Sandra, Elizabeth, Arlety y Susel, afortunada de haber compartido este viaje juntas, gracias por las noches de risa y llanto, por los largos estudios que terminaban en las mismas dudas, y los tantos momentos que nunca olvidaré, por poderlas llamar AMIGAS.

A mi querido vecino Reinier, gracias por tu amor, tu café, tu desorden y tus dudas a las 3 de la mañana.

A mi familia y demás amigos que de una forma u otra han contribuido a este trabajo.

Gracias Dios mío por permitirme llegar hasta aquí.

DEDICATORIA

A mi primer amor, mi maestro y mi amigo, Mi PAPÁ

y a los otros pedacitos de mi corazón, Mateo, Marcos, Verónica y el próximo que está por llegar.

RESUMEN

El objetivo de cualquier sistema de salud es mejorar la salud de donde se aplique, y una de las formas de lograr esto es la prevención de enfermedades. Por ello cobra especial importancia el estudio de la relación de las enfermedades con el espacio. Los Sistemas de Información Geográfica brindan la posibilidad de extraer conocimientos sobre las tendencias territoriales y su relación con los niveles de salud de determinada zona, sin embargo, los trabajos reportados en la literatura consultada no incluyen la componente espacial de los datos, lo que viola el principio de la primera ley de la geografía. Por otra parte, existe dispersión en las metodologías, herramientas y técnicas para abordar estudios de este tipo.

En esta investigación se presenta método no supervisado para la selección de rasgos en problemas de regionalización para su aplicación en estudios salubristas en la detección de fenómenos locales y globales. Se propone la utilización de autómatas celulares irregulares con aprendizaje para el proceso de selección de rasgos. La propuesta permite realizar estudios de regionalización según la primera ley de la geografía y garantiza la obtención de modelos más exactos. El método está conformado por cuatro etapas que cubren los procedimientos identificados en la literatura para este tipo de estudio. Las etapas propuestas se basan en el enfoque de análisis de datos geoespaciales y agrupamiento espacial, se denominan: Selección de capa y rasgos, Construcción y ponderado del grafo, Construcción del ICLA: inicialización y evolución, Generación de subconjuntos, soportado en una solución informática basada en software libre. Como parte de la validación se aplica el método en un caso de estudio disponible en la literatura.

Palabras claves: regionalización, selección de rasgos no supervisada, Sistema de Información Geográfica, estudios salubristas.

ABSTRACT

The goal of any health system is to improve health wherever it is applied, and one of the ways to achieve this is disease prevention. For this reason, the study of the relationship between diseases and space takes special importance. Geographic Information Systems offers the possibility of extracting knowledge about territorial trends and their relationship with the health levels of a certain area, however, the works reported in the consulted literature do not include the spatial component of the data, which violates the principle of the first law of geography. On the other hand, there is a dispersion in the methodologies, tools and techniques to approach studies of this type.

This research presents an unsupervised method for the selection of feautures in regionalization problems for its application in health studies in the detection of local and global phenomena. The use of learning irregular cellular automata is proposed for the feature selection process. The proposal allows for regionalization studies according to the first law of geography and guarantees the obtaining of more exact models. The method is made up of four stages that cover the procedures identified in the literature for this type of study. The proposed stages are based on the geospatial data analysis approach and spatial clustering, they are called: Selection of layer and features, Construction and weighting of the graph, Construction of the ICLA: initialization and evolution, Generation of subsets, supported in a computer-based solution in free software. As part of the validation, the method is applied in one case studies available in the literature.

Keywords: regionalization, unsupervised selection of features, Geographic Information System, health studies.

ÍNDICE GENERAL

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

INTRODUCCIÓN

Desde tiempos remotos el hombre ha podido establecer una relación entre el espacio y los problemas de salud. Cuando se observan a detalle los datos obtenidos a nivel mundial sobre una enfermedad específica, se encuentran diferencias entre las distintas regiones, registrándose en algunas cifras muy elevadas y en otras casi inexistentes. Estos resultados evidencian una relación directa entre los diferentes comportamientos territoriales y los problemas de salud (Dueñas Fernández 2016; Pérez Betancourt, González Polanco, Febles Rodríguez 2018).

El desarrollo socioeconómico permite la utilización del conocimiento adquirido a partir de las diferentes tendencias territoriales en el establecimiento de políticas de salud eficientes para la atención de una enfermedad o la distribución de recursos y servicios. En este sentido la Regionalización denota como una valiosa herramienta para analizar el comportamiento de variables en el espacio (Tenbensel 2016). Es considerada como un procedimiento técnico-administrativo de descentralización que consiste en dividir una región en porciones más pequeñas con un objetivo específico (Macleod, Cannon, Ko, Schade, Wright, Lin, Holt, Gore, Dash 2018). En los estudios salubristas, las técnicas de regionalización constituyen una herramienta poderosa debido a que permiten el análisis de la realidad territorial basándose en sus patrones y en las predicciones que pueden realizarse a partir de estos (Vali, Rashidian, Jalili, Omidvari, Jeddian 2017).

Los problemas de regionalización entran dentro de la clasificación de problemas de agrupamiento con restricciones espaciales, donde se podría pensar que lo más significativo es disponer de la máxima información posible. Por lo que puede parecer que cuanto mayor sea el número de atributos empleados mejor. Sin embargo la práctica ha demostrado que el rendimiento de los algoritmos de agrupamiento se deteriora ante la abundancia de información; debido a que muchos atributos pueden ser completamente irrelevantes para el problema o varios atributos redundantes pueden estar proporcionando la misma información (Chandrashekar, Sahin 2014). Por tanto, se hace imperante reducir el número de los datos sin perjudicar la resolución del problema, y mejorar los resultados en diversos aspectos (Lu, Chen, Yan, Jin, Xue, Gao 2017).

Las técnicas de reducción de datos son la respuesta a este problema, y dentro de estas la selección de atributos o rasgos reduce el tamaño de los datos eligiendo las variables más

influyentes en el problema, sin sacrificar la calidad del modelo. A pesar de su importancia para lograr un resultado óptimo, en la literatura consultada (Adams, Kanaroglou, Coulibaly 2016; Yuan, Tan, Cheruvelil, Collins, Soranno 2019; Kim, Dean, Kim, Chun 2016; Beauchemin 2019; Heřmanovský, Havlíček, Hanel, Pech 2017; Miranda 2017; Bianchi, Bruni, Reale, Sforzi 2016; Ayudiani, Akbar 2017; Brantley, Davis, Goodman, Callaghan, Barfield 2017; Welke, Pasquali, Lin, Backer, Overman, Romano, Karamlou 2020; Chhabra, Dimick 2016; França, McManus 2018; Morrone 2018; Lumpkin, Stitzenberg 2018) la utilización de técnicas de selección de rasgos en problemas de regionalización es prácticamente inexistente.

Debido a las potencialidades que brinda la información geográfica el uso de los Sistemas de Información Geográfica (SIG) ha aumentado considerablemente, fundamentado en las potencialidades que ofrecen para gestionar dicha información. Los SIG son herramientas básicas para la confección de mapas digitales y para los análisis geoespaciales, en todas las esferas del saber, que van más allá de análisis estadísticos y que tributan a una mejor planificación de infraestructura por ejemplo en: estudio demográfico, análisis de vías de transporte, distribución de recursos, distribución y comportamiento de enfermedades en salud (Liu, Wang, Wright, Cheng, Li, Liu 2017; Wu, Wang, Duan, Ouyang, Huang, Zuo 2016).

En nuestro país, específicamente en el sector de salud pública, los SIG se han utilizado principalmente en el análisis y distribución de los problemas de salud, fundamentalmente con enfoque a la estadística médica o en los modelos epidemiológicos. Su uso se ha enfocado principalmente en la representación de la distribución espacial de las enfermedades para mostrar geográficamente las tasas de incidencia con objetivos puramente descriptivos, también han sido utilizados para formular hipótesis relacionadas con la etiología de enfermedades y documentar o establecer el marco de estudios de la epidemiología (Rodríguez 2018; Cuéllar Luna, Gutiérrez Soto 2014). Estos elementos influyen en el análisis de la relación espacial de indicadores en diferentes áreas y en la capacidad de gestión de las entidades de salud. Las medidas de similitud empleadas consideran las características con igual importancia y están enfocadas a los datos temáticos. Este tratamiento deja de lado la componente espacial de los datos geográficos y por tanto dificulta la incorporación del espacio en el proceso, favoreciendo la aparición de regiones con territorios separados, incumpliendo con la primera Ley de la Geografía y que propicia hipótesis o modelos

inexactos e inconsistentes (González Polanco 2019; Delgado Acosta, González Moreno, Valdés Gómez, Hernández Malpica, Montenegro Calderón, Rodríguez Buergo 2015).
La incorporación de la componente espacial en los estudios salubristas se dificulta debido a las escasas y dispersas existencias de herramientas que cumplan esta labor, a el acceso limitado a los ISG por los costos que implican y que no existe un gran conocimiento de las herramientas ni personal adiestrado en el área y las tecnologías.
Debido a lo anterior la aplicación de técnicas de selección de rasgos en estudios salubristas basados en problemas de regionalización se hace inevitable ante la necesidad de obtener estudios más completos y relevantes que permitan aumentar el bienestar de la sociedad, formular nuevas teorías o modificar las existentes, solucionar problemas del individuo, la comunidad y el medio ambiente, crear nuevas tecnologías, facilitar la identificación de las ubicaciones geográficas de establecimientos de salud y grupos de población que presentan mayor riesgo de enfermar o de morir prematuramente y por tanto que requieran mayor atención preventiva, curativa o de promoción de salud.

A partir de la situación problemática descrita se define el siguiente **problema científico:** el insuficiente tratamiento a los rasgos en problemas de regionalización limita su aplicación en estudios salubristas para la detección de fenómenos locales y globales.

El **objeto de estudio:** Selección de rasgos

Para dar solución al problema se trazó el siguiente **objetivo general:** Desarrollar un método no supervisado para la selección de rasgos en problemas de regionalización que permita extender su campo de aplicación a la detección de fenómenos locales y globales.

Enmarcado en el **campo de acción:** Selección de rasgos en problemas de regionalización.

El objetivo general se desagrega en los siguientes **objetivos específicos:**

1. Construir el marco teórico referencial relacionado con la selección de rasgos.
2. Diseñar un algoritmo basado en autómatas celulares para la generación de subconjuntos de rasgos.

3. Diseñar un algoritmo para la selección de rasgos no supervisada en problemas de regionalización.

4. Implementar un plugin para el Sistema de Información Geográfica QGis a partir del método propuesto.

5. Verificar la solución informática propuesta aplicando pruebas de aceptación e índices de validación.

Se utilizan los siguientes **métodos de investigación:**

Métodos teóricos:

- **Análisis-síntesis:** para el estudio de las fuentes bibliográficas existentes relacionadas con el tema, identificando los elementos más importantes y necesarios para dar solución al problema planteado.
- **Histórico-lógico:** para el estudio crítico de los trabajos anteriores y utilizar estos como puntos de referencia y comparación de los resultados alcanzados.
- **Análisis documental:** con la consulta de la literatura especializada en las temáticas afines de la investigación.
- **Modelación:** para la representación explícita de la solución propuesta.

Métodos empíricos:

- **Cuasiexperimento:** para validar la propuesta se aplica el método de la presente investigación a dos casos de estudios disponibles en la literatura.
- **Observación:** se usó para adquirir información necesaria durante todas las fases de la investigación además de que permite ver desde diferentes puntos de vista la solución del problema.

CAPÍTULO 1. REFERENTES TEÓRICOS SOBRE REGIONALIZACIÓN Y SELECCIÓN DE RASGOS

En este capítulo se presentan los elementos que conforman los fundamentos teóricos relacionados con el objeto de estudio de la investigación. Se detallan los componentes y características de los SIG y se abordan definiciones de regionalización, su importancia y aplicaciones. Se explica el agrupamiento espacial, así como los conceptos de dependencia espacial, autocorrelación y su importancia para el tratamiento de los métodos de agrupamiento espacial con restricciones. Se detalla la selección de atributos y sus clasificaciones, específicamente los distintos enfoques de los métodos de selección de atributos no supervisados y los subtipos existentes entre ellos. Por último, se plantea la utilización de autómatas celulares irregulares como un nuevo acercamiento para la selección de atributos relevantes en problemas de regionalización.

1.1 Regionalización

La regionalización es el proceso de separar un espacio territorial en fracciones más pequeñas con un fin específico, atendiendo es a este fin varían tanto el proceso como la finalidad. Los procesos de regionalización permiten analizar el conocimiento del estado y de tendencias territoriales en múltiples disciplinas. Se realizan con un propósito determinado, por lo que pueden servir a un bien general o a requerimientos específicos sectoriales (Höwer, Oberst, Madlener 2019).

La regionalización se basa en el uso de información geográfica, el análisis espacial y el análisis estadístico multivariante, sus resultados resumen la realidad territorial en sus principales componentes, brindan patrones intrínsecos de la realidad socioeconómica del territorio, permiten realizar comparaciones entre los mismos y proporcionan un marco para la predicción y extrapolación basado en el comportamiento de las regiones obtenidas. Esta provee el marco espacial utilizado en varias disciplinas, entre las que se incluyen hidrología, ecología, ciencias políticas, geomorfología, ciencias medioambientales y economía, así como para aplicaciones públicas o administración de recursos naturales (He, Ling, Zhang, Gong 2018; Lebecherel, Andréassian, Perrin 2016; Salazar, Goldstein, Yang, Gause, Swarup, Hsiung, Rangel, Goldin, Abdullah 2016; Heřmanovský, Havlíček, Hanel, Pech 2017).

Este proceso delinea el paisaje geográfico en unidades contiguas espacialmente, conocidas como regiones o zonas, este proceso tiene como objetivo la creación de regiones homogéneas (landscape types en idioma inglés, LT) que comparten propiedades similares para una interpretación más entendible y fácil; se define de la siguiente manera: (Miranda 2017; He, Ling, Zhang, Gong 2018).

Se tiene $S = \{S_1, S_2, \dots, S_n\}$ que son todos los objetos espaciales situados en un territorio T, donde $S_i =$ es un vector de atributos d-dimensional, S_{i_j} es el valor del atributo del objeto espacial i en el atributo j y n es el número de objetos espaciales en S. Tal que: $i = 1na_i = Tya_i \cap a_j = \emptyset parai \neq j$

La regionalización de T es la partición de S en k regiones $\{G_1, G_2, \dots, G_k$ tal que: $i = 1nG_i = TyG_i \cap G_j = \emptyset parai \neq j; i, j \in \{1,2, \dots, k\}$.

Los métodos de regionalización multivariada se dividen en dos categorías, cualitativos y cuantitativos. Para los métodos cualitativos, los expertos separan las regiones con características del paisaje similares de múltiples mapas mediante interpretación visual; estos métodos no son exactos, debido a que la interpretación visual de un mapa especifico varía dependiendo del analista. Para los métodos cuantitativos, se utilizan enfoques de agrupamiento como el algoritmo K-means y los métodos de agrupamiento jerárquico para particionar el área geográfica en regiones más pequeñas.

A lo largo de los años se han desarrollado múltiples métodos de regionalización, que han dado solución a disimiles problemáticas, la ***Tabla 1-1*** muestra un recorrido por los diferentes métodos elaborados y los enfoques desde los que fueron construidos.

Método	***Temas Tratados***	***Estudio***
Regionalización mediante algoritmos de agrupamiento convencionales	Regionalización en dos etapas	(Openshaw 1973) (Openshaw, Blake, Wymer 1995)
	Sensibilidad de los resultados de regionalización de acuerdo a diferentes algoritmos de agrupamiento	(Johnston 1968) (Lankford 1969) (Fischer 1980)
	Consideración de límites naturales	(Mills 1967) (Segal, Weinberger 1977) (George, Lamar, Wallace 1997)
	Consideración de agregaciones preexistentes	(George, Lamar, Wallace 1997)

Regionalización mediante maximización de compacidad regional	Asignaciones fraccionarias	(Helbig, Orr, Roediger 1972)
	Uso de algoritmos genéticos	(Bação, Lobo, Painho 2005)
Modelos exactos de optimización	Formas alternativas de satisfacer la restricción de contigüidad	(Duque, Artís, Ramos 2006) (Macmillan, Pierce 1994)
	Restricción de contigüidad basada en el poder de la matriz de contigüidad	(Garfinkel, Nemhauser 1970)
	Solución obtenida por la unión de regiones viables prediseñadas	(Mehrotra, Johnson, Nemhauser 1998)
Adaptación de algoritmos de agrupamiento jerárquico	Comparación de métodos jerárquicos	(Spence 1968) (Webster, Burrough 1972) (Byfuglien, Nordgård 1973) (Margules, Faith, Belbin 1985)
	Métodos jerárquicos e identificación de patrones espaciales.	(Byfuglien, Nordgård 1973)
	Contigüidad espacial basada en el límite de contigüidad	(Perruchet 1983)
Modificación de una solución inicial factible	Uso de recocido simulado	(Browdy 1990) (Macmillan, Pierce 1994) (Macmillan 2001) (Openshaw, Rao 1995)
	Formas de verificar la contigüidad espacial	(Macmillan, Pierce 1994) (Macmillan 2001)
Modelos basados en teoría de grafos	Uso de árbol de expansión	(Maravalle, Simeone 1995)
Mezcla de modelos heurísticos	División del problema de regionalización en subproblemas	(Duque 2004)
	Uso de concentración heurística	(Duque, Church 2004)
	Uso de destilación heurística	(Middleton 2006)

Tabla 1-1*: Métodos de Regionalización.*

Existen otros métodos de regionalización desarrollados como la familia de algoritmos REDCAP, esta es un conjunto de métodos basado en arboles de expansión mínimo. Primero construyen un árbol de contigüidad espacial mediante enfoques de agrupamiento jerárquico y luego eliminan los bordes del árbol para generar regiones. Existen tres tipos de enfoques de agrupamiento jerárquico utilizados en estos métodos; enlace completo (CLK), enlace promedio (ALK) y enlace único (SLK). Para cada estrategia de agrupamiento existen dos restricciones de contigüidad espacial diferentes: restricciones de primer orden y restricciones

de orden completo, lo cual genera seis métodos diferentes de regionalización: Primer orden CLK, ALK, SLK y Orden completo CLK, ALK y SLK (He, Ling, Zhang, Gong 2018).

Otro método propuesto en (Miranda 2017) es el RegK-Means, el cual plantea una adaptación del algoritmo K-means que utiliza la relación de vecindad especificada entre los objetos como una restricción en la definición de grupos.

En (Bianchi, Bruni, Reale, Sforzi 2016) los autores plantean un nuevo enfoque para problemas de regionalización, específicamente para la identificación de Áreas locales de mercado laboral (LLMA). Lo convierten en un problema de partición de grafo y la solución es obtenida resolviendo una secuencia de problemas de corte mínimo sobre un grafo no dirigido de las interacciones entre las localidades.

Estos son algunos ejemplos de métodos de regionalización que se pueden encontrar en la literatura, aunque estos métodos pertenecientes al enfoque cuantitativo reducen la subjetividad del enfoque cualitativo proveyendo una forma sistemática y reproducible de identificar regiones, tienen varias limitaciones, relacionadas con el hecho de que se apoyan en técnicas no supervisadas. La primera, la determinación del número ¨real¨ de tipos de paisajes es problemática. Existen herramientas automáticas que pueden ser utilizadas de guía, por ejemplo, el índice Davies-Bouldin y el criterio de silueta. Sin embargo, estas herramientas no están específicamente diseñadas para capturar la percepción humana de LT. Segundo, no existe garantía de que el algoritmo de agrupamiento (clustering en idioma inglés) capturará la compleja noción de homogeneidad percibida por los humanos para definir regiones homogéneas. Sin embargo, la mayor dificultad de los métodos de regionalización es la restricción de contigüidad espacial, ya que la regionalización puede considerarse como un tipo de problema de agrupamiento con la restricción de que todos los objetos espaciales en cada región son espacialmente contiguos; lo que se conoce como agrupamiento espacial con restricciones.

1.2 Sistemas de Información Geográfica (SIG)

Las características concretas de la información geográfica hacen que sea necesario el desarrollo de herramientas altamente especializadas para su gestión. Estas herramientas son

los Sistemas de Información Geográfica (SIG), cuyo diseño y concepción permite recoger toda la riqueza de matices de esta información y, aún más, permiten rentabilizar dicha información. Un Sistema de Información Geográfica (SIG) define un conjunto de procedimientos con capacidad de construir modelos o representaciones del mundo real a partir de datos geográficos de localización cierta y mensurable (Rodríguez 2018; Fletcher, Caprarelli 2016).

Los SIG pueden ser usados en cualquier aplicación cuyo objetivo principal sea gestionar algún tipo de información georreferenciada; referida a los elementos o fenómenos que tienen lugar sobre la superficie terrestre (Katayama, Yokoyama, Yako-Suketomo, Okamoto, Tango, Inaba 2014). Son una herramienta especializada esencial para poder manipular con eficacia la información geográfica porque aumentan su accesibilidad, su exactitud y, en general, garantizan la eficacia de los resultados de las decisiones a tomar.

El desarrollo y la diversificación de los Sistemas de Información Geográfica (SIG) posibilita que actualmente esta poderosa herramienta sea aplicable en un campo tan sensible e importante como la salud pública, teniendo en cuenta fundamentalmente la gran cantidad de datos que se genera en esta actividad y que estos puedan ser representados para gestionar dicha información mediante la combinación de datos demográficos (edad, sexo, distribución), con datos de salud (tipos de enfermedades, incidencias, prevalencia, características clínicas o patológicas), características del medio natural (clima, altitud, precipitación) y cualquier otra información que el especialista considere necesaria. Se pueden obtener resultados tales como precisar las áreas de influencia de determinada enfermedad, la ocurrencia por edades, sexo o por determinadas condiciones del medio ambiente natural y la posibilidad de que se presente en otras áreas por tener las mismas condiciones naturales o demográficas (Abousaeidi, Fauzi, Muhamad 2016; Campbell, Shin 2018; Huang, Ma, Xiao, Sun 2019). Las aplicaciones de los SIG en la salud pública**,** como la gestión de recursos sanitarios, logística o análisis de enfermedades para la optimización de recursos humanos y estudios epidemiológicos es una de las temáticas menos conocidas pero más extendidas dentro de los Sistemas de Información Geográficas. La base de todo SIG permite la representación de la información a través de mapas. Este es uno de los terrenos más básicos dentro de la sanidad y la salud. Sencillas representaciones de datos como la distribución de

enfermedades o la localización de centros de salud pudieran ser algunos de los ejemplos. Gracias a visores cartográficos y representaciones interactivas, los datos sanitarios pasan de ser algo fijo a mostrar información dinámica, tanto informativa como divulgativamente (Roberto 2016; Beiranvand, Karimi, Delpishehd, Sayehmiri, Soleimani, Ghalavandi 2016; Ahmadi, Ramazani, Rezagholi, Yavari 2018; Wang 2020).

En la literatura consultada se encuentra la utilización de diferentes SIG, como por ejemplo: gvSIG, ArcView, MapInfo y QGis (Rodríguez 2018; Delgado Acosta, González Moreno, Valdés Gómez, Hernández Malpica, Montenegro Calderón, Rodríguez Buergo 2015). En esta investigación se utiliza QGis, destacándose por su licencia GNU. Es una aplicación escritorio, con un entorno sencillo, amigable, muy intuitivo y fácil de utilizar. Además, es multiplataforma, posibilita conexión a base de datos PostgreSQL y PostGIS. Permite la incorporación de nuevos módulos y funcionalidades implementadas en C++ y Python, manipula formatos ráster y vectoriales a través de las bibliotecas GDAL y OGR, así como bases de datos.

Los SIG en estudios salubristas se han utilizado principalmente para procesar la estadística médica y en investigaciones epidemiológicas que estudian la magnitud y distribución de distintos problemas sanitarios en las poblaciones, así como en la vigilancia, análisis, monitoreo, evaluación de intervenciones, gestión y toma de decisiones vinculadas con este campo. En su accionar en esta esfera, los SIG se vinculan con otras disciplinas tales como la epidemiología, la geografía, la bioestadística y la tecnología de la información. Debido a la creciente acumulación de información espacial producto del desarrollo de los sistemas informáticos y en especial de los SIG, se ha propiciado la aplicación de técnicas de minería de datos espaciales para dar soporte a la toma de decisiones y para estudios de diagnóstico-intervención-evaluación en la salud (González Polanco 2019).

1.3 Agrupamiento espacial

La revolución digital ha hecho posible que la información digitalizada sea fácil de capturar, procesar, almacenar, distribuir y transmitir. Con el progreso de la informática, de las tecnologías relacionadas y la expansión de su uso en diferentes aspectos de la vida ha aumentado la recopilación y el almacenamiento de gran cantidad de datos. La minería de

datos (MD) es el intento humano de encontrarle sentido a la gran cantidad de datos que actualmente puede ser almacenada (B, Siddiqui, Arain 2019; Atluri, Karpatne, Kumar 2018).

La minería reúne ventajas de diversos campos como lo son: la estadística, la inteligencia artificial, la computación gráfica, las redes neuronales, entre otros. La MD brinda la posibilidad de extraer información de grandes volúmenes de datos que por separados no resultan de utilidad, esta información puede ser en forma de patrones, cambios, asociaciones y estructuras que permiten la creación de modelos que sustentan la toma de decisiones (Kavakiotis, Tsave, Salifoglou, Maglaveras, Vlahavas, Chouvarda 2017).

Actualmente la información tiene un papel importante en la competitividad y la productividad de cualquier organización. Entre los diversos tipos de datos que se manejan, la información de tipo geográfico está tomando gran relevancia en la toma de decisiones organizacionales. A partir de esto han surgido herramientas, métodos y técnicas para extraer conocimiento de este tipo de datos: la minería de datos espacial (Requia, Koutrakis, Roig, Adams, Santos 2016; Aksac, Özyer, Alhajj 2019).

Los datos de tipo espacial reflejan la primera ley de la geografía enunciada por (Tobler 1979), que afirma que en el análisis geográfico todo está relacionado con todo, pero las cosas cercanas están más relacionadas entre sí que las cosas lejanas. Se entiende entonces por dato espacial todo aquel que tiene asociada una referencia geográfica, de tal modo que podemos localizar exactamente dónde sucede dentro de un mapa (Moise, Ruiz 2016; Shaweno, Karmakar, Alene, Ragonnet, Clements, Trauer, Denholm, McBryde 2018; Bi, Azman, Satter, Khan, Ahmed, Riaj, Gurley, Lessler 2016).

El objetivo de la minería de datos espaciales es encontrar relaciones entre objetos de tipo espacial y no espacial a través de relaciones, como las topológicas, las de orientación espacial y las de distancia de información. La extracción de patrones en conjuntos de datos espaciales es más complicada que en conjuntos de datos tradicionales (datos numéricos), debido a la complejidad de los datos, relaciones y autocorrelación espacial.

La dependencia espacial se refiere a la relación entre los datos georreferenciados debido a la naturaleza de la variable bajo estudio y el tamaño, forma y configuración de las unidades espaciales. Cuanto menores son las unidades espaciales, mayor será la probabilidad de que

las unidades cercanas sean espacialmente dependientes. Si las unidades son largas y estrechas, las posibilidades de dependencia espacial con unidades cercanas serán mayores que si las unidades son más compactas. En otras palabras, la dependencia espacial de un objeto se puede definir como el simple hecho de estar presente, lo que implica que tiene un lugar y una forma, considerando sus relaciones, causas y consecuencias. Al ser así, puede encontrarse sujeto a la acción de un agente, existe una relación entre ambos, y ésta puede ser considerada como una amenaza. La dependencia espacial de dicho objeto con los elementos a su alrededor se analiza a partir de las relaciones espaciales existentes entre ellos. Estas relaciones espaciales son conceptos que surgen de la interacción entre el espacio y los eventos que en él ocurren, así como todas sus combinaciones (Aturinde, Farnaghi, Pilesjö, Mansourian 2019; Liang, Chen, Chien, Chen 2018; Zhang, Chen, Li, Taft, Yao, Bai, Xing 2019). Estos tipos de relaciones se organizan con base en la mayor o menor dominancia, sea de las propiedades del espacio o de las propiedades de los eventos. En este contexto los tres grupos de relaciones espaciales son: relaciones métricas, relaciones topológicas y relaciones de organización.

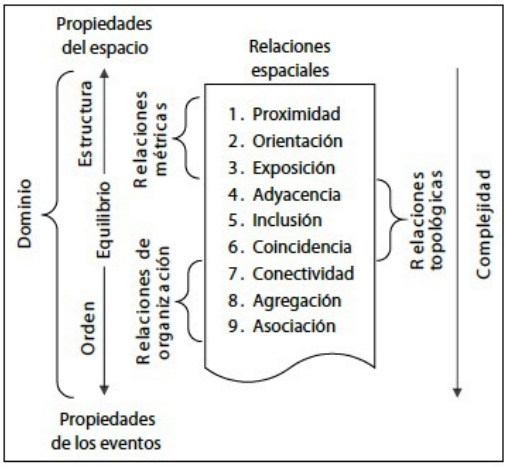


Figura 1-1: *Relaciones Espaciales, tomado de (Pérez 2014; Universidad Nacional Autónoma de México 2015).*

La principal característica de la minería de datos espaciales es que los objetos espaciales incluyen restricciones de contigüidad, por tanto, además de la similitud de valor, la similitud espacial también se tiene en cuenta para el análisis espacial. El término autocorrelación espacial hace referencia a lo planteado en la ley de Tobler, es decir, a la existencia de una correlación de la variable consigo misma, de tal modo que los valores de esta variable en un

punto guardan relación directa con los de esa misma variable en otros puntos cercanos. Por ejemplo: supóngase que se estudian una serie de poblaciones cercanas en las cuales se mide el porcentaje de personas afectadas por una determinada enfermedad infecciosa. Cabe esperar que, puesto que los habitantes de esas poblaciones están relacionados entre sí de diversas formas, la distribución de los valores recogidos obedezca en parte a la existencia de dichas relaciones. Por ejemplo, si en una población contraen la enfermedad un número dado de habitantes, es más factible que estos puedan contagiar a los de las poblaciones cercanas que a los de otros núcleos más alejados (Alene, Clements 2019; Cárceles-Álvarez, Ortega-García, López-Hernández, Orozco-Llamas, Espinosa-López, Tobarra-Sánchez, Alvarez 2017; Huang, Tam, Chern, Lung, Chen, Wu 2018).

Por lo anterior, es probable que alrededor de una población con muchos casos de la enfermedad haya otras también con un elevado número de afectados, mientras que una población con pocos casos esté rodeada de otras también con escasa afección. En el caso de la enfermedad infecciosa, los valores altos suelen tener en su entorno valores también altos, y de modo similar sucede para valores bajos, entonces existe una autocorrelación espacial positiva. Puede, no obstante, existir una autocorrelación espacial negativa, si los valores altos se rodean de valores bajos y viceversa (Olaya 2016; Last, Burr, Alexander, Harding-Esch, Roberts, Nabicassa, Cassama, Mabey, Holland, Bailey 2017; Alene, Clements 2019).

En caso de no existir ningún tipo de autocorrelación espacial, se tiene que los datos recogidos en una serie de puntos son independientes entre sí y no se afectan mutuamente, sin que tenga influencia de la distancia.

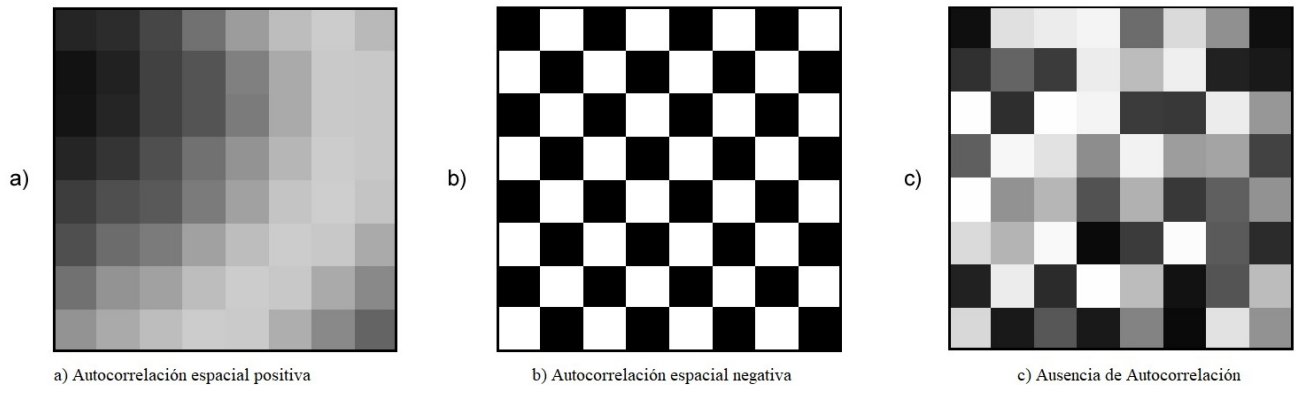


Figura 1-2: *Autocorrelación espacial, tomado de (Olaya 2016).*

El análisis de autocorrelación espacial se realiza para verificar si los datos tienden a ser dispersos (autocorrelación negativa), aleatorios o agrupados (autocorrelación positiva). Las consecuencias de la existencia de autocorrelación espacial son numerosas y de gran importancia. Por una parte, muchos de los análisis estadísticos suponen la independencia de la variable. Puesto que existe una dependencia de la componente espacial, es necesario para obtener resultados correctos introducir dicha componente espacial como una variable más. Puede también sacarse provecho de la existencia de una dependencia espacial, puesto que los puntos cercanos a uno dado guardan relación con este y pueden emplearse para estimar su valor.

Los métodos de minería espacial se clasifican en cinco grupos:

Figura 1-3*: Métodos para el descubrimiento de conocimiento en base de datos espaciales tomado de (Palacio 2002).*

De acuerdo a lo mencionado anteriormente de que la regionalización puede considerarse como un caso específico de problema de agrupamiento con la restricción de contigüidad espacial se destaca el uso de los métodos de agrupamiento espacial para los problemas de regionalización sobre los otros enfoques.

Los métodos de agrupamiento espacial se dividen en tres grupos fundamentales: particionales, jerárquicos y basados en localidad.

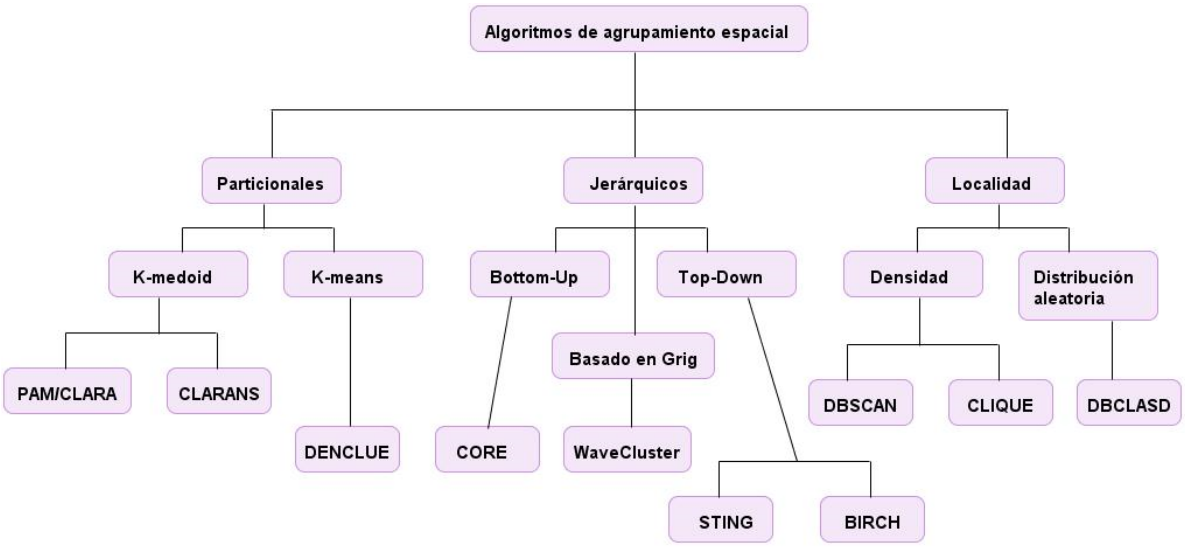


Figura 1-4: *Algoritmos de agrupamiento, tomado de (Peña Suárez 2017)*

Las técnicas de minería de datos que extraen modelos a partir de ejemplos tienden a obtener modelos complejos conforme crece el volumen de datos del conjunto sobre el cual se aplican. Entre los principales temas a considerar en las tareas de agrupamiento de datos espaciales tenemos la determinación de la validez de un grupo, la estimación de la tasa de error del sistema de reconocimiento, el conocimiento previo que se tenga sobre el dominio de datos, las propiedades del algoritmo de agrupamiento y la dimensionalidad de los datos (Smith, Auala, Tambo, Haindongo, Katokele, Uusiku, Gosling, Kleinschmidt, Mumbengegwi, Sturrock 2017; Yamaoka, Suzuki, Inoue, Ishikawa, Tango 2020). En lo referente a los problemas de regionalización este último se convierte en un obstáculo debido a la gran información que guardan los datos de tipo geográfico. Por tanto, se hace necesario la utilización de técnicas de reducción de datos con el fin de poder decidir qué datos deben ser utilizados para el análisis permitiendo eliminar información irrelevante y agilizar el proceso de agrupamiento.

1.4 Selección de rasgos

Las técnicas de reducción de datos permiten reducir el elevado tamaño de los conjuntos de datos, evitando inconvenientes adicionales como el aumento del tiempo de respuesta de los

modelos, la sensibilidad al ruido y la posibilidad de sobreajuste de los modelos. Al emplear un mayor número de datos, aumenta la probabilidad de retener ejemplos ruidosos provocando que esos ejemplos de escasa calidad afecten los modelos, modificando la adecuada clasificación. La obtención de modelos de gran tamaño conlleva a que la solución sea poco comprensible para la mente humana, se hace muy engorroso comprender la solución de un problema que emplea cientos de ejemplos o reglas para representarla, por tanto, cuanto menor sea su tamaño, más comprensible será (Solorio-Fernández, Carrasco-Ochoa, Martínez-Trinidad 2018; Abualigah, Khader, Hanandeh 2018; Remeseiro, Bolon-Canedo 2019)

Por tanto, es necesario un preprocesamiento de los datos en el que se disminuya el tamaño del conjunto almacenado, la ***Figura 1-5*** muestra las diferentes técnicas que se pueden emplear para llevar a cabo la reducción de datos.

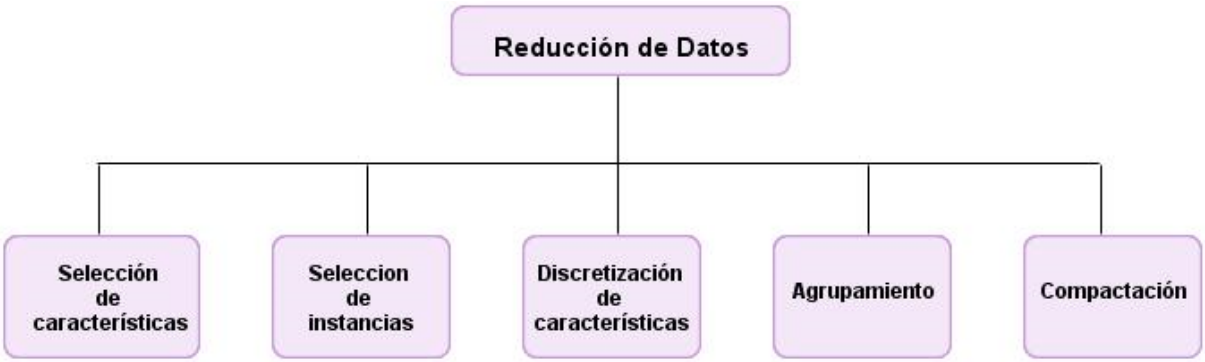


Figura 1-5: *Técnicas de reducción de datos, tomado de (Herrera 2006).*

La selección de características, también conocida como selección de atributos o rasgos es una técnica usada habitualmente en la minería de datos que describe el proceso mediante el cual las entradas de datos se reducen a un tamaño apropiado para su procesamiento y análisis (Violini 2014; Miao, Niu 2016; Sheikhpour, Sarram, Gharaghani, Chahooki 2017). La selección de rasgos aparece en diferentes áreas como el reconocimiento de patrones, aprendizaje automático, la minería de datos y el análisis estadístico. En todas estas áreas, los objetos estudiados incluyen en su descripción atributos irrelevantes y redundantes, lo que puede afectar significativamente el análisis de los datos, resultando en sesgos o modelos incorrectos. La selección de atributos es ampliamente utilizada en tareas como clasificación,

regresión o agrupamiento. Esto no solo reduce la dimensionalidad de los datos; facilitando su visualización y entendimiento; sino también conduce a la generación de modelos más compactos con mejor capacidad de generalización (He, Beauseroy, Smolarz 2015; Luo, Nie, Chang, Yang, Hauptmann, Zheng 2018; Mafarja, Aljarah, Faris, Hammouri, Al-Zoubi, Mirjalili 2019; Li, Liu 2017).

De acuerdo a la información disponible en los conjuntos de datos, los métodos de selección de atributos pueden ser clasificados como supervisados, semi-supervisados y no supervisados. Los métodos supervisados requieren un conjunto de datos etiquetados para identificar y seleccionar atributos relevantes; esta etiqueta, asignada a cada objeto en el conjunto, puede ser una categoría, un valor ordenado o un valor real, dependiendo de la tarea específica. Los métodos semi-supervisados solamente requieren que algunos objetos estén etiquetados. Por último, los métodos de selección de atributos no supervisados (UFS por sus siglas en inglés) no requieren conjuntos de datos supervisados o etiquetados (Solorio-Fernández, Carrasco-Ochoa, Martínez-Trinidad 2020).

1.4.1 Selección de rasgos no supervisada (UFS)

En las últimas décadas, se han propuesto muchos métodos de selección de atributos, la gran mayoría desarrollados para tareas de clasificación supervisada. Sin embargo, debido al desarrollo tecnológico en los últimos años, así como a la cantidad de datos sin etiquetar generada en diferentes aplicaciones como la minería de texto, bioinformática, redes sociales y detección de intrusos, por mencionar algunos; los métodos UFS han ganado un interés significativo en la comunidad científica (Zhu, Zhu, Hu, Zhang, Zuo 2017; Zheng, Zhu, Wen, Zhu, Yu, Gan 2020; Nie, Zhu, Li 2016). Además, los métodos UFS tienen dos ventajas importantes, la primera, son imparciales y funcionan bien cuando no existe conocimiento previo, y la segunda, reducen el riesgo de sobreajuste de datos, en contraste con los métodos supervisados que pueden ser incapaces de manejar una nueva clase de datos (Solorio-Fernández, Carrasco-Ochoa, Martínez-Trinidad 2018; Zhu, Xu, Hu, Zhang 2018; Hu, Zhu, Cheng, He, Yan, Song, Zhang 2017).

Un problema de selección de características se define formalmente como (Solorio Fernández, Carrasco Ochoa, Martínez Trinidad 2020):

Sea X el conujnto original de atributos, donde $|X| = m > 0$. Sea la medida de evaluación $L: P(X) \rightarrow R^{+\cup\{0\}}$; la cual es la medida a maximizar, donde $P(X)$ denota el conjunto de las partes de X, esto es, $P(X) = \{S \vee S \subset P\}$. Se denota $X_k \subseteq X$ como un subconjunto de atributos seleccionados, con $|X_k| = k$. Por tanto $X_0 = \emptyset y X_m = X$.

Sea L la medida de evaluación a optimizar (maximizar). La selección de un subconjunto de atributos puede hacerse bajo dos premisas:

- Sea k tal que $0 < k < m$. Encontrar $X_k \subset X$ tal que $L(X_m)$ sea máximo.
- Sea un valor real $L_{mín} > 0$,valor minimo de L que va a ser aceptado. Encontrar $X_k \subseteq X$ con el menor k tal que $L(X_k) \geq L_{mín}$.

En estas condiciones, siempre existirá un subconjunto óptimo de atributos, no necesariamente único.

Los algoritmos de selección de atributos realizan una búsqueda en el espacio de subconjuntos de atributos. Por lo que de modo general los algoritmos deben tener:

Punto de comienzo: se establece un punto o puntos de comienzo en el espacio de búsqueda lo cual establece una dirección: si se inicia la búsqueda por un conjunto vacío, la única dirección posible es hacia delante; este procedimiento se denomina forward selection por su nombre en inglés. Si se comienza por el conjunto completo se procede a ir eliminando atributos: esto se denomina backward selection. También pueden plantearse combinaciones o variantes de estos modelos básicos.

Organización de la búsqueda: el espacio de búsqueda se puede recorrer siguiendo diferentes criterios. Los tres tipos fundamentales son: heurística, exhaustiva y aleatoria.

Medida de evaluación: criterio por el cual se mide la bondad del conjunto de atributos seleccionado. Algunos ejemplos son la ganancia de información, la distancia, la dependencia, la consistencia y la precisión.

Criterio de parada: criterio para dejar de buscar en el espacio de posibilidades. Se puede parar cuando no se mejore el mérito alcanzado, normalmente se sigue buscando mientras la precisión no se degrade o finalmente encuentre el otro extremo del espacio de búsqueda.

Con lo planteado anteriormente se puede definir un proceso clásico para cualquier algoritmo de selección de características:

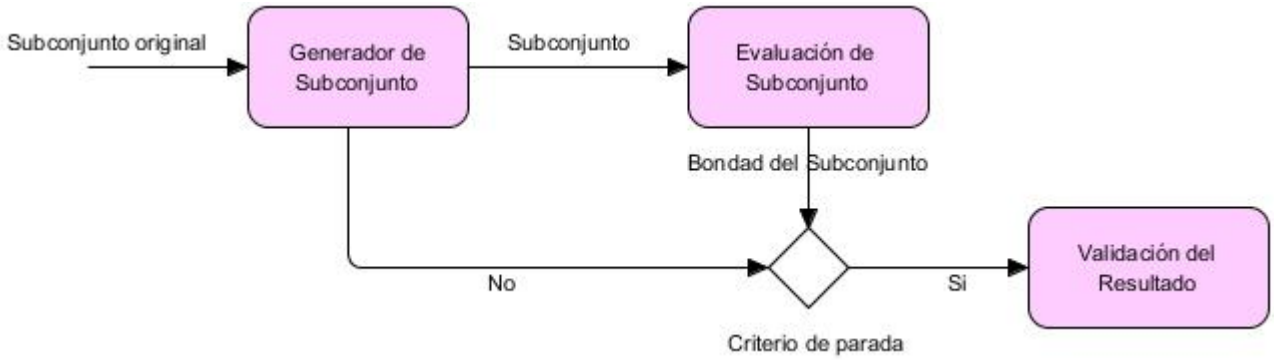


Figura 1-6: *Proceso de selección de características, tomado de (Chandrashekar, Sahin 2014).*

De acuerdo a la estrategia utilizada para la selección de atributos, los UFS se dividen en tres enfoques principales, y estos a su vez se subdividen en categorías de acuerdo al procedimiento que utilizan para la selección.

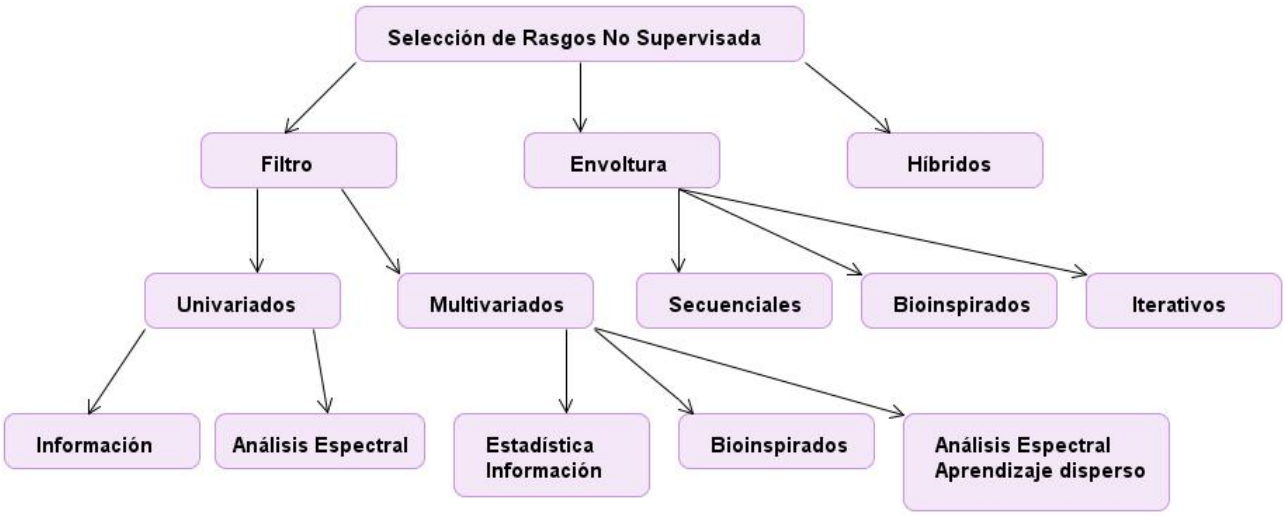


Figura 1-7: *Taxonomía de los métodos UFS, tomado de (Solorio-Fernández, Carrasco-Ochoa, Martínez-Trinidad 2018).*

1. Los métodos de filtro seleccionan las características relevantes a través de los datos en sí, es decir, los atributos se evalúan en función de las propiedades intrínsecas de los datos, sin utilizar ningún algoritmo de clustering que pueda guiar la búsqueda de

atributos relevantes. La principal característica de los métodos de filtro es su rapidez y escalabilidad.

- **Univariados:** son UFS basados en ranking que utilizan criterios para evaluar cada atributo obteniendo una lista ordenada de donde se escoge un subconjunto de atributos finales. Estos métodos identifican y eliminan atributos irrelevantes, pero no eliminan atributos redundantes ya que no tienen en cuenta las dependencias entre atributos.

    - ❖ **Información:** los métodos basados en información evalúan la relevancia de cada atributo en función de la Teoría de la Información. Evalúan el grado de dispersión de los datos a través de medidas como la entropía y la divergencia para identificar estructuras de clúster en los datos.
    - ❖ **Análisis Espectral:** siguen la idea de modelar o identificar la estructura de los datos local o global utilizando el sistema de Laplacian derivado de una matriz de similitud.

- **Multivariados:** siguen el mismo procedimiento de los univariados, con la diferencia de que estos pueden manejar atributos irrelevantes y redundantes, obteniendo mejores resultados.
    - ❖ **Estadística/ Información:** realizan la selección utilizando teoría estadística y o de información, medidas como varianza-covarianza, correlación lineal, entropía, información mutua, entre otros.
    - ❖ **Bioinspirados:** usan estrategias estocásticas de búsqueda basadas en el paradigma de inteligencia de enjambres para encontrar un subconjunto de rasgos que satisfaga algún criterio de calidad.
    - ❖ **Aprendizaje Disperso/ Análisis Espectral:** la selección de características se logra como parte del proceso de aprendizaje, comúnmente a través de la optimización de un modelo de regresión restringido.

2. Los métodos de envoltura evalúan los subconjuntos de atributos utilizando los resultados de un algoritmo especifico de clustering. Estos métodos se caracterizan por

encontrar subconjuntos de atributos que contribuyen a mejorar la calidad de los resultados del algoritmo de clustering usado para la selección. La mayor desventaja de estos métodos es que usualmente tienen un gran costo computacional y están limitados a utilizarse en conjunto con un algoritmo de clustering particular.

- **Secuenciales:** en estos métodos los atributos son añadidos y removidos secuencialmente. Los métodos basados en búsqueda secuencial son fáciles de implementar y rápidos.
- **Bioinspirados:** estos métodos intentan incorporar la aleatoriedad en el proceso de búsqueda, con el objetivo de escapar de los óptimos locales.
- **Iterativos:** abordan el problema de selección de características no supervisadas considerándolo un problema de estimación y, por lo tanto, evitando una búsqueda combinatoria.

3. Los métodos híbridos tratan de explotar las cualidades de los métodos de filtro y los de envoltura, tratando de llevar a un compromiso entre eficiencia (costo computacional) y eficacia (calidad en la tarea objetivo al usar los atributos seleccionados). Existen dos grupos de métodos híbridos, los métodos basados en ranking y los no-basados en ranking.

Enfoque	***Ventajas***	***Desventajas***
Filtro	Rápido Escalable Independiente del algoritmo de clustering	Ignora la interacción con algoritmos de agrupamiento
Envoltura	Interactúa con el algoritmo de clustering utilizado Modela dependencias entre rasgos	Riesgo de sobreajuste Gran costo computacional La selección es específica para el algoritmo de agrupamiento utilizado
Híbridos	Interactúa con el algoritmo de clustering utilizado Modela dependencias entre rasgos Consume menos tiempo que los de envoltura	La selección es específica para el algoritmo de agrupamiento utilizado

Tabla 1-2: *Ventajas y desventajas generales de los métodos UFS con respecto a su enfoque, tomado de (Solorio Fernández, Carrasco Ochoa, Martínez Trinidad 2020).*

En la literatura consultada podemos encontrar diferentes métodos de selección de rasgos todos con diferentes enfoques (Lu, Chen, Yan, Jin, Xue, Gao 2017; Abualigah, Khader,

Hanandeh 2018; Remeseiro, Bolon-Canedo 2019; Miao, Niu 2016; Luo, Nie, Chang, Yang, Hauptmann, Zheng 2018; Li, Guo, Liu, Liu 2019; Zhu, Xu, Hu, Zhang 2018; Hu, Zhu, Cheng, He, Yan, Song, Zhang 2017; Solorio-Fernández, Carrasco-Ochoa, Martínez-Trinidad 2018; Zhu, Zhu, Hu, Zhang, Zuo 2017; Zheng, Zhu, Wen, Zhu, Yu, Gan 2020; Nie, Zhu, Li 2016), sin embargo, debido a la complejidad de los datos espaciales, en este trabajo se plantea una nueva propuesta basada en la teoría de autómatas celulares para el proceso de selección de rasgos debido a su utilidad en la modelación de sistemas dinámicos.

1.5 Teoría de Autómatas celulares

Para el estudio de la matemática es fundamental el uso y desarrollo de herramientas que expliquen los fenómenos del entorno. Esto se logra través de modelos matemáticos que den respuesta a dichos fenómenos. Lo cual ha permitido un gran avance en el estudio del caos y de los sistemas dinámicos. Por otro lado, la computación puede verse como la transformación de la información, donde al inicio de este proceso siempre hay condiciones iniciales. Sin embargo, existen procesos de cómputo donde nuevas entradas de información surgen durante el proceso mismo. Esta información nueva a veces determina el resultado del proceso, lo que implica un enfoque distinto para el estudio de la computación, donde el sistema sea capaz de cambiar de comportamiento ante cualquier perturbación, incorporando información nueva durante el proceso. Para auxiliar a ambos enfoques, es de mucha ayuda el estudio y simulación de sistemas dinámicos, evitando las desventajas existentes en la matemática clásica para expresar la complejidad de estos sistemas. Debido a esto surge un método de modelización conocido como autómatas celulares (Cano Rojas, Rojas Matas 2016; Omrani, Tayyebi, Pijanowski 2017).

Los autómatas celulares (AC) surgen en la década de 1940 con John Von Neumann, que intentaba modelar una máquina que fuera capaz de autorreplicarse, llegando así a un modelo matemático de dicha máquina con reglas complicadas sobre una red rectangular. Inicialmente fueron interpretados como conjuntos de células que crecían, se reproducían y morían a medida que pasaba el tiempo. A esta similitud con el crecimiento de las células se le debe su nombre.

Un autómata celular es un modelo matemático para un sistema dinámico, compuesto por un conjunto de celdas o células que adquieren distintos estados o valores. Estos estados son alterados de un instante a otro en unidades de tiempo discreto, es decir, que se puede cuantificar con valores enteros a intervalos regulares. De esta manera este conjunto de células logra una evolución según una determinada expresión matemática, que es sensible a los estados de las células vecinas, lo cual se conoce como regla de transición local.

El aspecto que más caracteriza a los AC es su capacidad de lograr una serie de propiedades que surgen de la propia dinámica local a través del paso del tiempo y no desde un inicio, aplicándose a todo el sistema en general. Por lo tanto, analizar las propiedades globales de un AC desde su comienzo, complejo por naturaleza, se torna difícil; a no ser por vía de la simulación, partiendo de un estado o configuración inicial de células y cambiando en cada instante los estados de todas ellas de forma síncrona (Rezapoor Mirsaleh, Meybodi 2016).

La definición de un AC requiere mencionar sus elementos básicos:

- **Arreglo Regular:** es el espacio de evoluciones, ya sea un plano de dos dimensiones o un espacio n-dimensional, donde cada división homogénea de arreglo se llama célula.
- **Conjunto de Estados:** conjunto finito de donde cada elemento o célula del arreglo toma un valor. También se denomina alfabeto y puede ser expresado en valores o colores.
- **Configuración Inicial:** asignación de un estado a cada una de las células del espacio de evolución inicial del sistema.
- **Vecindades:** conjunto contiguo de células y posición relativa respecto a cada una de ellas. A cada vecindad diferente corresponde un elemento del conjunto de estados.
- **Función Local:** regla de evolución que determina el comportamiento del AC. Se conforma de una célula central y sus vecindades. Define como debe cambiar de estado cada célula dependiendo de los estados anteriores de sus vecindades. Puede ser una expresión algebraica o un grupo de ecuaciones.

Los autómatas celulares son utilizados con éxito en distintas disciplinas. Por ejemplo, en la Física es una de las técnicas más interesantes para simular fenómenos concretos en dinámica

de fluidos. En Biología los AC representan desde mediados de los años 80 una seria alternativa a la modelización con ecuaciones diferenciales en el estudio de los sistemas complejos. Uno de los factores que más ha contribuido a su uso es la sencillez con que se pueden realizar simulaciones. A finales de los años 90 el uso de los AC abarcó numerosas disciplinas, siendo de gran utilidad en el estudio de sistemas biológicos: reproducción, autoorganización y evolución. Una de las aplicaciones más interesantes hoy en día, es en las Ciencias de la Computación, donde los AC han permitido a los investigadores construir modelos para el estudio del procesamiento de información en paralelo, así como el diseño de computadoras cuya arquitectura sea basada en principios y materiales biológicos (Fernández Fraga 2014).

1.5.1 Autómatas Celulares con Aprendizaje

Un autómata celular con aprendizaje (CLA) (Mason, Gu 1986) es una combinación de autómata celular (CA) (Packard, Wolfram 1985) y autómata de aprendizaje (LA) (Narendra, Thathachar 2012). La idea básica del CLA es usar el autómata de aprendizaje para ajustar la probabilidad del estado de transición de un autómata celular estocástico. Este modelo, que abre un nuevo paradigma de aprendizaje, es superior a los AC debido a su capacidad de aprender y también es superior a los autómatas de aprendizaje porque consiste en una colección de LA que interactúan entre sí (Vafashoar, Meybodi 2019; Rezvanian, Moradabadi 2019).

Un CLA es un CA en el que se asigna un número de LA a cada celda. Cada LA que reside en una celda particular determina su acción (estado) en función de su vector de probabilidad de acción.

Se define un autómata celular con aprendizaje como una estructura $A = (Z^d, N, \varphi, A, F)$ donde (Rezvanian, Moradabadi 2019):

• Z^d: presenta la red de d-tuplas de números enteros.

• $N = \{\overline{x}_1, \overline{x}_2, \dots, \overline{x}_m\}$ es un subconjunto finito de Z^d que se llama vector de vecindario, donde $\overline{x}_1 \in Z^d$.

• φ: denota el conjunto finito de estados. φ_i representa el estado de la célula c_i.

• A: es el conjunto de LA que reside en las células del CLA.

• $F^i: \varphi_i \rightarrow \beta$: define la regla local del CLA para cada celda c_i, donde β es el conjunto de valores posibles para la señal de refuerzo y calcula el refuerzo para cada LA utilizando las acciones elegidas de los LAs vecinos.

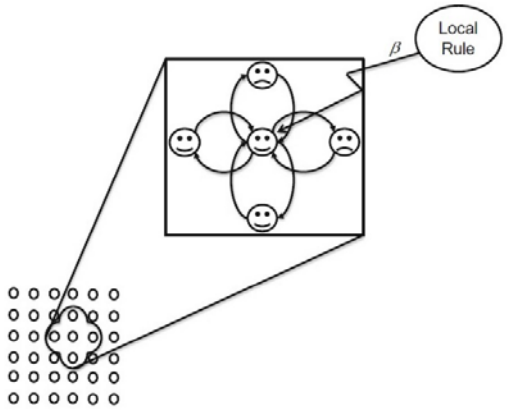


Figura 1-8: *Operación de un CLA, tomado de (Rezvanian, Moradabadi 2019).*

1.5.2 Autómatas Celulares Irregulares con Aprendizaje (ICLA)

El autómata celular irregular con aprendizaje (Irregular Cellular Learning Automaton por su nombre en inglés) (Esnaashari, Meybodi 2008) es una generalización del CLA tradicional que elimina la limitación de la estructura matricial. Esto se hace necesario para aplicaciones relacionadas con gráficos, redes sociales, redes inalámbricas, de sensores y sistemas de redes inmunes, que no se pueden modelar con una estructura de matriz (Esnaashari, Meybodi 2010; Rezapoor Mirsaleh, Meybodi 2016).

Se considera un ICLA como un grafo no dirigido en el que cada nodo es una celda que está equipada con un autómata de aprendizaje, y los nodos vecinos de cualquier nodo particular constituyen el entorno local de esa celda. El LA que reside en una celda particular determina su estado (acción) de acuerdo con su vector de probabilidad de acción. El entorno local de un LA no es estacionario porque la probabilidad de acción los vectores de los LAs vecinos varían durante la evolución del ICLA.

El funcionamiento del ICLA es similar al funcionamiento del CLA. En el primer paso el estado interno de cada celda se especifica sobre la base del vector de probabilidad de acción del LA que reside en esa celda. En el segundo paso, la regla del ICLA determina la señal de refuerzo del LA que reside en cada celda. Finalmente, cada LA actualiza su vector de

probabilidad de acción sobre la base de la señal de refuerzo suministrada y el estado interno de la célula. Este proceso continúa hasta que se obtiene el resultado deseado (Ghavipour, Meybodi 2017; Esnaashari, Meybodi 2018).

Se define un autómata celular irregular con aprendizaje como una estructura $A = (G\langle E, V\rangle, F, A, F)$ donde (Rezvanian, Moradabadi 2019):

•G: es un gráfico no dirigido, con V como el conjunto de vértices (celdas) y E como el conjunto de aristas (relaciones de adyacencia).

• F: denota el conjunto finito de estados. F_i representa el estado de la célula c_i.

• A: es el conjunto de LA cada uno de los cuales está asignado a una celda de la ICLA.

• $F^i: \varphi_i \rightarrow \beta$: define la regla local del CLA para cada celda c_i, donde $\varphi_i = \{\varphi_i | \{i, j\} \in E\} \cup \{\varphi_i\}$ es el conjunto de estados de todos los vecinos de c_i y β es el conjunto de valores que la señal de refuerzo puede tomar.

Nota: en la formalización del ICLA, no se da una definición explícita para la vecindad de cada celda, ya que está implícitamente definida en la definición del gráfico G.

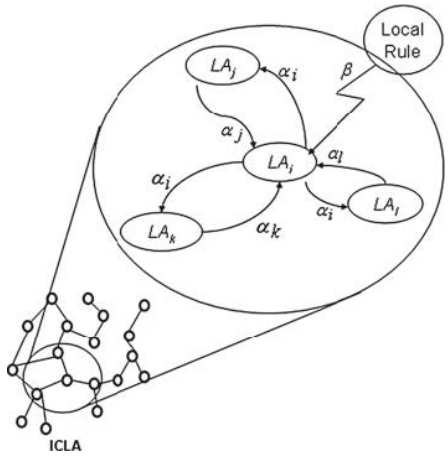


Figura 1-9: *Irregular Cellular Learning Automaton, tomado de (Rezvanian, Moradabadi 2019).*

1.6 Herramientas, Lenguajes y Tecnologías a utilizar

En todo proceso investigativo es necesaria la utilización de sistemas que faciliten la tarea de reducir los datos obtenidos durante dicho proceso, estos sistemas permiten automatizar, agilizar y organizar los trabajos que se generan durante el transcurso de la investigación. A

continuación, se describen las herramientas, tecnologías y lenguajes que se van a utilizar para la realización de este trabajo.

1.6.1 Lenguaje de Modelado

UML es el acrónimo de Lenguaje Unificado de Modelado, este es la lengua franca del desarrollo de software, permitiendo especificar, visualizar y documentar los artefactos de un sistema, incluida su estructura y diseño, utilizándose para el modelado del negocio y sistemas de software (*Unified Modeling Language* 2019). Ofrece un estándar para describir los modelos, incluyendo aspectos conceptuales tales como procesos, funciones del sistema, y aspectos concretos como expresiones de lenguajes de programación, esquemas de bases de datos y componentes reutilizables.

1.6.2 Herramienta CASE *Computer Aided Software Engineering*

Las herramientas CASE son diversas aplicaciones informáticas destinadas a aumentar la productividad en el desarrollo de software reduciendo el coste de las mismas en términos de tiempo y de dinero. Estas herramientas ofrecen soporte en todos los aspectos del ciclo de vida de desarrollo del software en tareas como el proceso de realizar un diseño del proyecto, cálculo de costes, implementación de parte del código automáticamente con el diseño dado, compilación automática, documentación o detección de errores (Beltrán 2018).

Visual Paradigm

Es una herramienta UML profesional que soporta el ciclo de vida completo del desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue, permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación. (Mendoza Peña 2016).

Las principales características de la herramienta son:

- Soporta las últimas versiones del UML.
- Posee un poderoso generador de documentación y reportes en formato PDF, HTTP y JPG.
- Proporciona soporte para varios lenguajes en la generación de código e ingeniería inversa como: Java, C++, CORBA IDL, PHP, Ada y Python.

- Disponibilidad en múltiples plataformas (Windows, Linux)
- Capacidades de ingeniería directa e inversa.

Se selecciona Visual Paradigm para UML en su versión 8.1 como herramienta para el modelado UML, pues permite trabajar de forma colaborativa, hacer un trabajo organizado y ágil. Posibilita la realización de los diagramas necesarios para el desarrollo y mejor entendimiento de la aplicación. Al ser seleccionado el lenguaje de modelado UML, es conveniente tener en cuenta su vinculación con Visual Paradigm, resaltando que este último presenta abundante documentación y demostraciones interactivas.

1.6.3 Lenguaje de programación

Los lenguajes de programación son un conjunto de símbolos junto a un conjunto de reglas sintácticas y semánticas que definen su estructura y el significado de sus elementos y expresiones. Constan de un léxico, una sintaxis y una semántica (Gé Vaillant 2020). Existen muchos lenguajes para el desarrollo de aplicaciones, surgidos a partir de las tendencias y necesidades de los escenarios. El análisis se centró fundamentalmente en el lenguaje Python a partir de la posibilidad que brinda QGis para integrar componentes implementados en este lenguaje.

Python

Es un lenguaje de programación potente y fácil. Tiene estructuras de datos eficientes de alto nivel y un enfoque simple pero efectivo para la programación orientada a objetos. La elegante sintaxis y la escritura dinámica de Python, junto con su naturaleza interpretada, lo convierten en un lenguaje ideal para la creación de scripts y el desarrollo rápido de aplicaciones en muchas áreas en la mayoría de las plataformas.

Se seleccionó Python en su versión 3.7 porque al seleccionar QGis como el software que soportará la

integración de la solución, el lenguaje de programación más eficiente y conveniente para utilizar

es Python. Su sintaxis es clara, simple y sencilla logrando de esta manera que los programas elaborados en este lenguaje parezcan pseudocódigo. Además, el tipado dinámico, el gestor

de memoria, la gran cantidad de librerías disponibles y la potencia del lenguaje, entre otros, hacen que desarrollar una aplicación en Python sea sencillo y rápido.

PyQT

PyQt es un conjunto de enlaces Python para la biblioteca gráfica Qt. El módulo está desarrollado por la firma británica Riverbank Computing y se encuentra disponible para Windows, GNU/Linux y Mac OS bajo diferentes licencias. PyQt posee un número importantes de herramientas que gestionan su manipulación y posibilita adecuarse a las distintas plataformas de software.

Utilizando PyQt en su versión 5.0 en el desarrollo de la herramienta informática, se puede crear una interfaz visual sencilla y sin muchos contratiempos, ya que PyQt posee los componentes visuales necesarios para su desarrollo, así como una abundante documentación y ejemplos.

QT Designer

Qt Designer es una herramienta que permite acelerar el desarrollo de interfaces multilenguaje debido a que genera un archivo XML cuyo contenido es el formato de dicha interfaz, pudiéndolo convertir con los programas pertinentes a cada lenguaje. Esta herramienta provee características muy poderosas como la previa visualización de la interfaz, soporte para widgets y un editor de propiedades con gran variedad de opciones.

En correspondencia con la elección anterior de PyQt, se ha decidido emplear Qt Designer en su versión 5.0 como elemento que soportará el diseño de las interfaces. Su utilización permite la creación de las interfaces visuales de la aplicación de forma sencilla, además de la fácil manipulación de las variables de configuración de cada una de ellas.

1.6.4 Entorno de desarrollo integrado

Un entorno de desarrollo integrado (IDE, por sus siglas en inglés) es una herramienta que permite a los desarrolladores de software escribir sus programas en uno o más lenguajes.

Consiste básicamente en una plataforma en la que se integran un editor de código, un compilador, un depurador y una interfaz gráfica de usuario.

Pycharm

Pycharm es un editor de código inteligente que proporciona soporte de primera clase para los lenguajes de programación: Python, JavaScript, CoffeeScript, TypeScript, HTML/CSS, Cython, lenguajes de plantilla, AngularJS y Node.js, y otros menos utilizados. Pycharm funciona en las plataformas Windows, Mac OS y Linux con una única clave de licencia, también ofrece un espacio de trabajo con colores personalizables y atajos de teclado.

Se seleccionó como IDE, Pycharm en su versión 3.2, ya que ofrece comprobación de errores sobre la marcha, auto-completación inteligente de código, fácil navegación en el proyecto y soluciones rápidas. Pycharm mantiene la calidad del código bajo control con asistencia a pruebas, chequeos, refactorizaciones inteligentes y una serie de inspecciones, lo que ayuda a escribir un código limpio y fácil de mantener (Nafiul Islam 2015).

1.6.5 Quantum Gis (QGis)

Quantum Gis (o QGis) es un Sistema de Información Geográfica (SIG) tipo escritorio, muy intuitivo y fácil de utilizar que pretende ofrecer a usuarios con necesidades básicas un entorno sencillo y agradable. Su licencia es GNU, y por tanto se trata de código libre. Es multiplataforma y se pueden encontrar versiones para diferentes sistemas operativos: GNU/Linux, Unix, Mac OS y Microsoft Windows.

Salió oficialmente como producto de la fundación OSGeo en 2008. Permite manipular formatos ráster y vectoriales a través de las bibliotecas GDAL y OGR, así como bases de datos. Hasta no hace mucho, era uno de los pocos editores de PostGIS para la plataforma Windows y se destaca por su sencillez y velocidad (Rodríguez 2018).

1.6.6 Gestor de Base de Datos

Los Gestores de Bases de Datos (GBD) son herramientas que permiten el almacenamiento, manipulación y consulta de datos pertenecientes a una base de datos organizada en uno o varios ficheros. Los GBD son la herramienta más adecuada para almacenar los datos en un sistema de información debido a sus características de seguridad, recuperación ante fallos,

gestión centralizada, estandarización del lenguaje de consulta y funcionalidad avanzada. Además, actúan como interfaz entre los programas de aplicación y el sistema operativo. El objetivo principal es proporcionar un entorno eficiente a la hora de almacenar y recuperar la información de las bases de datos. Estos softwares facilitan el proceso de definir, construir y manipular bases de datos para diversas aplicaciones (Olaya 2016).

PostgreSQL

PostgreSQL es un sistema de GBD objeto-relacional, de propósito general, multiusuario y de código abierto desarrollado en el Departamento de Informática de la Universidad de California en Berkeley, que soporta gran parte del estándar SQL. Ofrece modernas características como consultas complejas, disparadores, vistas, integridad transaccional, control de concurrencia multiversión. Puede ser extendido por el usuario añadiendo tipos de datos, operadores, funciones agregadas, funciones ventanas y funciones recursivas, métodos de indexado y lenguajes procedurales (Ordóñez, Ríos, Castillo 2017).

Fue seleccionado PostgreSQL en su versión 9.0, teniendo en cuenta que es un GBD multiplataforma y de código abierto. Además, se valoró la existencia de la extensión PostGIS para permitir el trabajo con datos espaciales.

PostGIS

Para añadir soporte a PostgreSQL de objetos geográficos se utilizó la herramienta PostGIS en su versión 2.1.5. Este módulo convierte la base de datos objeto-relacional PostgreSQL en una base de datos espacial para su utilización en SIG. PostGIS incluye un conjunto de operaciones para realizar consultas espaciales muy bien optimizadas por sus índices R-Tree y su integración con el planificador de consultas de PostgreSQL. Utiliza las librerías Proj4 para dar soporte a la transformación dinámica de coordenadas y la biblioteca GEOS para realizar operaciones de geometría. Utiliza bloqueo a nivel de fila, permitiendo a múltiples procesos trabajar con las tablas espaciales concurrentemente y asegurando la integridad de los datos (Corti, Kraft, Mather, Park 2014).

PgAdmin

Como aplicación gráfica para gestionar el GBD PostgreSQL se utilizó la herramienta PgAdmin III en su versión 1.20.0. PgAdmin está diseñado para responder a las necesidades de todos los usuarios, desde escribir consultas SQL simples hasta desarrollar bases de datos complejas. Soporta todas las características de PostgreSQL y facilita enormemente la administración. La aplicación también incluye un editor SQL con resaltado de sintaxis, un editor de código para la parte del servidor y un agente para lanzar scripts programados. La conexión al servidor puede hacerse mediante conexión TCP/IP o Unix Domain Sockets (en plataformas Unix), y puede encriptarse mediante SSL para mayor seguridad (Ordóñez, Ríos, Castillo 2017).

1.7 Metodología de desarrollo

El desarrollo de software no es una tarea sencilla, por mucho tiempo esta labor se llevó a cabo sin una metodología definida. Al respecto algunos autores definen una metodología como una colección de procedimientos, técnicas, herramientas y documentos auxiliares que ayudan a los desarrolladores de software en sus esfuerzos por implementar nuevos sistemas de información. En las dos últimas décadas se ha entablado un intenso debate entre dos grandes corrientes. Por un lado, las denominadas metodologías tradicionales, centradas en el control del proceso, con un riguroso seguimiento de las actividades involucradas en ellas y por otro, las metodologías ágiles, centradas en el factor humano, en la colaboración y participación del cliente en el proceso de desarrollo y a un incesante incremento de software con iteraciones muy cortas(Kruchten, Fraser, Coallier 2019; Suryantara, Andry 2018).

Programación Extrema

Programación extrema, XP por sus siglas en inglés se basa en una serie de reglas y principios que se han ido gestando a lo largo de toda la historia de la ingeniería del software. Usadas conjuntamente proporcionan una nueva metodología de desarrollo de software que se puede englobar dentro de las metodologías ligeras y se clasifica como evolutiva (Campos, Martínez 2015).

Esta es una metodología ágil centrada en potenciar las relaciones interpersonales como clave para el éxito en desarrollo de software, promoviendo el trabajo en equipo, preocupándose por el aprendizaje de los desarrolladores, y propiciando un buen clima de trabajo. XP se basa en

retroalimentación continua entre el cliente y el equipo de desarrollo, comunicación fluida entre todos los participantes, simplicidad en las soluciones implementadas y coraje para enfrentar los cambios. XP se define como especialmente adecuada para proyectos con requisitos imprecisos y muy cambiantes, y donde existe un alto riesgo técnico (Kruchten, Fraser, Coallier 2019; Sadath, Karim, Gill 2018).

Características de la metodología XP:

- XP es una metodología "liviana" que no tiene en cuenta la utilización de casos de uso y la generación de una extensa documentación.
- XP tiene asociado un ciclo de vida y es considerado a su vez un proceso.
- La tendencia de entregar software en espacios de tiempo cada vez más pequeños con exigencias de costos reducidos y altos estándares de calidad.
- XP define Historias de Usuario (HU) como base del software a desarrollar, estas historias las escribe el cliente y describen escenarios sobre el funcionamiento del programa. A partir de las HU y de la arquitectura perseguida se crea un plan de liberaciones entre el equipo de desarrollo y el cliente.

Fases de la metodología XP (Gómez, Duarte, Guevara 2014):

- **Planificación:** durante esta etapa se lleva a cabo el proceso de identificación y confección de las HU.
- **Diseño:** durante esta etapa se crea un diseño evolutivo que va mejorando incrementalmente y que permite hacer entregas pequeñas y frecuentes de valor para el cliente, basado principalmente en el desarrollo de las tarjetas Clase-Responsabilidad- Colaboración (CRC).
- **Desarrollo:** en esta fase se realiza la implementación de las HU que fueron seleccionadas por cada iteración. Al inicio se lleva a cabo un chequeo del plan de iteraciones por si es necesario realizar modificaciones. Como parte de este plan se crean tareas de ingeniería para ayudar a organizar la implementación exitosa de las HU.
- **Pruebas:** esta fase permite aumentar la seguridad de evitar efectos colaterales no deseados a la hora de realizar modificaciones y refactorizaciones. XP divide las

pruebas del sistema en dos grupos: pruebas unitarias, encargadas de verificar el código y diseñadas por los programadores, y pruebas de aceptación o pruebas funcionales destinadas a evaluar si al final de una iteración se consiguió la funcionalidad requerida diseñada por el cliente final.

El ciclo de desarrollo consiste en los siguientes pasos:

1. El cliente define el valor de negocio a implementar.
2. El programador estima el esfuerzo necesario para su implementación.
3. El cliente selecciona qué construir, de acuerdo con sus prioridades y las restricciones de tiempo.
4. El programador construye ese valor de negocio.
5. Vuelve al paso 1.

A partir de la investigación realizada de XP, se concluye que esta metodología responde a las necesidades principales de tiempo, entorno y cantidad de programadores, e incluye al cliente como parte fundamental del equipo de desarrollo. Además, se preocupa más en el avance exitoso del producto que en generar una documentación detallada del mismo, siendo capaz de adaptarse a los cambios de requisitos en cualquier punto del ciclo de vida del proyecto, por lo cual se escoge para el desarrollo de la propuesta.

1.8 Conclusiones del Capítulo

La construcción del marco teórico referencial de la investigación, relacionado con la regionalización de territorios generó las siguientes conclusiones:

- La definición del marco teórico referencial de la investigación relacionado con el proceso de regionalización de territorios, fundamentó la necesidad de implementar algoritmos de selección de rasgos para favorecer la creación de modelos más compactos en el proceso de clustering.
- Con el desarrollo de este capítulo se obtuvo un mejor dimensionamiento del problema a partir del análisis de los principales conceptos asociados a su solución. El análisis de los elementos básicos, así como técnicas de selección de rasgos y la teoría de

autómatas celulares, permitió conocer las características que poseen para darle solución al problema planteado.

- El uso de los Sistemas de Información Geográfica ha aumentado considerablemente, sin embargo, su utilización en el sector de la salud aún se limita a la visualización de mapas y no se explotan en su totalidad la componente espacial de los datos.
- La revisión del panorama actual de los softwares SIG condujo a la selección de QGis como el más adecuado para la integración de la propuesta.
- A partir del análisis de las herramientas y tecnologías se seleccionaron un conjunto de ellas, basadas en licencia de software libre, para obtener un producto de alta independencia tecnológica y multiplataforma. Se escogió la metodología XP para guiar el proceso de desarrollo de la solución.

CAPÍTULO 2. MÉTODO NO SUPERVISADO PARA LA SELECCIÓN DE RASGOS EN PROBLEMAS DE REGIONALIZACIÓN

En este capítulo se describe y fundamenta un método no supervisado para la selección de rasgos en problemas de regionalización que permita extender su campo de aplicación a la detección de fenómenos locales y globales.

Se presenta el paradigma empleado para ejecutar la investigación. Se especifican los requisitos de software y se obtienen los artefactos correspondientes a las fases de planificación, diseño e implementación de la metodología seleccionada. Se define la arquitectura y los principales patrones de diseño utilizados en el desarrollo de la solución. Se detallan las tareas de ingenierías que conforman cada HU definida en la fase de planificación y se establece el estándar de codificación que se estará utilizando en el desarrollo de la solución.

2.1 Paradigma utilizado para el diseño de la propuesta.

La presente investigación está enmarcada en la disciplina de los sistemas de información (SI). Los sistemas de información se implementan dentro de una organización con el fin de

mejorar la efectividad y eficiencia de esa organización, estos son esencialmente artefactos del conocimiento que capturan y representan este recurso en ciertos dominios. Dos paradigmas caracterizan gran parte de la investigación en la disciplina de Sistemas de Información: la ciencia del comportamiento y la ciencia del diseño. El primer paradigma busca el desarrollo y verificación de teorías que expliquen o pronostiquen el comportamiento humano u organizacional. Por su parte las ciencias del diseño tienen sus raíces en ingeniería y ciencias de lo artificial (Rodríguez, María 2019; Hatchuel, Le Masson, Reich, Subrahmanian 2018). Es fundamentalmente un paradigma de solución de problemas. Busca crear innovaciones que definan las ideas, prácticas, capacidades técnicas, y productos a través de los cuales el análisis, diseño, implementación, gestión y uso de los sistemas de información puedan ser logrados efectiva y eficientemente (Dresch, Lacerda, Antunes 2015; Venable, Pries-Heje, Baskerville 2016).

A partir de la naturaleza del problema abordado en esta investigación y la relación que existe entre su campo de acción y la disciplina de los SI, la propuesta se desarrolla bajo el paradigma de las ciencias del diseño. Desde este enfoque se debe producir un artefacto viable en la forma de un constructo, un modelo, un método o una instanciación.

Un constructo constituye el vocabulario conceptual de un dominio a partir del cual se pueden definir y comunicar el problema en cuestión y su solución. Los modelos representan el problema y solución a partir de un conjunto de proposiciones o sentencias que expresan relaciones entre constructos. Los métodos proveen guías sobre cómo resolver problemas y encontrar las soluciones, pueden expresarse a partir de algoritmos, descripciones textuales del proceso de búsqueda de las soluciones o combinaciones de ambas. Finalmente, las instanciaciones muestran la viabilidad de implementación de constructos, modelos y métodos a partir de su operacionalización, lo que facilita la evaluación concreta del artefacto que se instancia (González Polanco 2019).

Los constructos disponibles dentro del objeto de estudio que se aborda en la presente investigación son suficientes para describir adecuadamente el problema que se aborda. Además, los enfoques aportados en investigaciones precedentes con relación a estudios de regionalización son válidos, aunque diversos y no hacen uso de técnicas de selección de rasgos para reducir el volumen de datos. Por lo que se hace necesario introducir una propuesta que integre los enfoques disponibles para explotar el espacio de solución teniendo en cuenta

la componente espacial y la selección de rasgos. Se propone un método no supervisado para la selección de rasgos en problemas de regionalización que combina teoría de autómatas celulares irregulares con aprendizaje con el objetivo de obtener modelos más compactos. Se implementa una instanciación del método propuesto para evaluar su viabilidad.

Se planificó y ejecutó la investigación a partir del proceso definido para investigaciones bajo el paradigma de las ciencias del diseño y que define las etapas: identificación del problema y motivación, objetivos de la solución, diseño y desarrollo, demostración, evaluación y comunicación. Los elementos de las dos primeras etapas fueron presentados en la introducción de este documento. Las etapas de diseño y desarrollo abarcan la creación de los artefactos asociados a la solución e incluye la sistematización del vocabulario conceptual disponible y la identificación de los requisitos. El vocabulario conceptual y los requisitos fueron determinados a partir de los referentes analizados en el capítulo 1. Los artefactos creados serán descritos en los epígrafes siguientes. La demostración y evaluación de su desempeño en la solución del problema serán analizados en el capítulo 3. Como evidencias de la comunicación a la comunidad científica, los principales aportes de este trabajo serán publicados en artículos de revistas y presentados en diferentes conferencias científicas.

2.2 Método no supervisado para la selección de rasgos en problemas de regionalización para la detección de fenómenos locales y globales.

El aporte fundamental de esta investigación es un método no supervisado para la selección de rasgos en problemas de regionalización para la detección de fenómenos locales y globales basados en ICLA y SIG que da continuidad al desarrollo de herramientas y técnicas para el análisis espacial en estudios salubristas. El método está conformado por cuatro etapas que cubren los procedimientos identificados en la literatura para este tipo de estudio. Las etapas propuestas se basan en el enfoque de análisis de datos geoespaciales y se denominan: Selección de capa y rasgos, Construcción y ponderado del grafo, Construcción del ICLA: inicialización y evolución y Generación de subconjuntos. El objetivo de este método es el tratamiento de selección de rasgos en problemas de regionalización para la detección de fenómenos locales y globales como soporte metodológico a la toma de decisiones en estudios salubristas.

Se sustenta en los siguientes principios:

- **Integración** de técnicas de selección de rasgos, teoría de AC, regionalización y SIG para darle tratamiento a la componente espacial en la detección de fenómenos locales y globales en estudios salubristas.
- **Modelación** de la información entorno a la regionalización, selección de rasgos, los datos geoespaciales y los estudios salubristas.
- **Reutilización** de buenas prácticas relacionadas con estudios de regionalización y selección de rasgos como base para el desarrollo del método y su realización mediante analíticas de datos, que favorezca la incorporación de la espacialidad a la detección de fenómenos locales y globales.

Los enfoques de la propuesta son:

- **Holístico** con el estudio de los rasgos, el espacio en su conjunto y su complejidad, se identifican interacciones, particularidades y procesos que por lo regular no se perciben si se estudian los rasgos por separados y luego se llevan a la cartografía.
- **Estratégico** con la generación de un subconjunto óptimo que permita la regionalización de territorios para la detección de fenómenos locales y globales que facilite el establecimiento de objetivos claros a largo plazo y su conjunto de acciones a corto plazo para dar respuesta a las oportunidades y amenazas que impone el entorno, así como las fortalezas y debilidades.

Las cualidades que distinguen al método:

- **Integración:** el método técnicas de selección de rasgos, teoría de AC, regionalización y SIG para darle tratamiento a la componente espacial en la detección de fenómenos locales y globales en estudios salubristas. También se distingue por la integración de técnicas de análisis de datos geoespaciales en una solución informática que sirve de soporte tecnológico.

- **Usabilidad:** el enfoque de guía para la regionalización y la interfaz de la instanciación facilitan la integración de la cartografía y rasgos en los estudios sin necesidad de mucho dominio en este campo.
- **Fiabilidad:** la información que brinda se corresponde con el análisis de los indicadores aportados.
- **Flexibilidad:** a partir del uso de indicadores de naturaleza variada y un marco de trabajo para la regionalización se facilita adaptarse a cambios que se deseen incluir en los estudios.

Las premisas:

- Voluntad de las organizaciones y entidades administrativas de la salud pública para su utilización a diferentes niveles administrativos.
- Personal calificado para aplicarlo con rigor científico.
- Disponibilidad de la base cartográfica con la división político administrativa y la información asociada a los indicadores que se incluirán en el estudio.

La aplicabilidad del método se basa en la concepción de que puede ser empleado a diferentes niveles de dirección territorial y de administración de la salud. Puede ser extensible a estudios de regionalización en otros contextos donde se tengan identificados indicadores y las entidades administrativas asociadas.

Las entradas del método son:

- **Base cartográfica:** está formada por capas vectoriales que pueden ser de puntos, líneas o polígonos y responden a indicadores geoespaciales.
- **Rasgos:** conjunto de todos los rasgos de los objetos.

La salida del método permite la realización del proceso de regionalización con un menor volumen de datos, creando modelos más compactos y con mayor exactitud lo cual servirá de soporte para la toma de decisiones, aportando elementos asociados a distribuciones y procesos espaciales útiles para la definición de objetivos y planes en el tratamiento a problemas de salud.

Las salidas del método son:

- **Subconjunto de rasgos:** subconjunto óptimo de rasgos para la regionalización.

La ***Figura 2-1*** muestra una representación del método propuesto y sus características son descritas en los siguientes epígrafes del presente capítulo.

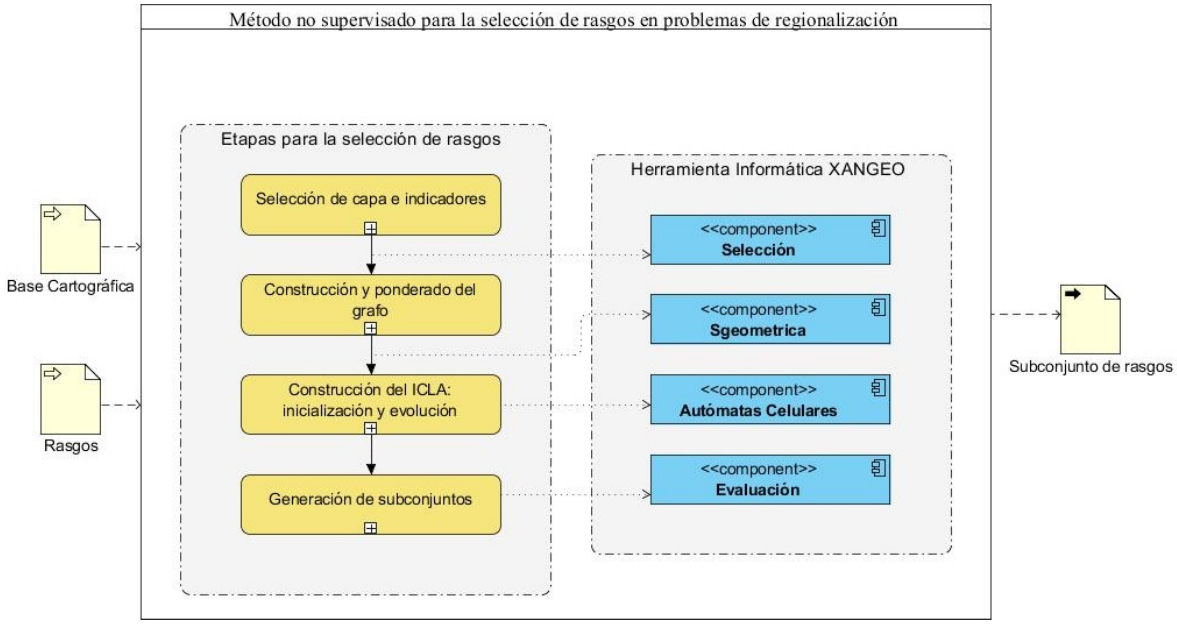


Figura 2-1: *Método no supervisado para la selección de rasgos en problemas de regionalización, elaboración propia.*

El método de selección de rasgos para regionalización propuesto está compuesto por un conjunto de etapas con sus procedimientos y el sistema informático XANGEO cuenta con los componentes necesarios para la ejecución del mencionado método. Posee un enfoque diferente a los estudios reportados en la literatura, al proponer la integración de indicadores geoespaciales y teoría de ICLA en el estudio.

El método que se propone abarca los procedimientos y etapas de los estudios reportados sobre regionalización desde un enfoque de la minería de datos geoespaciales. En él se incluyen cuatro etapas destinadas a la selección de los rasgos y capa base para el estudio, la construcción y ponderado del grafo, construcción del ICLA, su inicialización y evolución y la generación de subconjuntos. Todas estas etapas serán formalizadas en el epígrafe 2.3.

El sistema informático XANGEO contiene una instanciación del método propuesto e incluye la implementación de las técnicas de minería de datos geoespaciales identificadas en el **¡Error! No se encuentra el origen de la referencia.**. También propone un marco de integración entre el SIG QGis y las principales bibliotecas geoespaciales para su utilización

en estudios de regionalización. Su arquitectura, principales funcionalidades y componentes son descritos en el epígrafe **¡Error! No se encuentra el origen de la referencia.**.

2.3 Descripción de las etapas que conforman el Método no supervisado de selección de rasgos.

La etapa de selección de capa e indicadores tiene como objetivo obtener todos los rasgos y los territorios que serán objeto de estudio, por lo que es necesario disponer de un mapa vectorial e indicadores disponibles en fuentes estadísticas, o recopilados por el investigador utilizando diferentes técnicas y herramientas. La capa debe ser de polígonos que representan a los territorios objetos de estudio. Los datos temáticos corresponden a la información de otras fuentes como son datos de la población, factores de riesgos e indicadores de salud, por solo mencionar algunos.

Con los rasgos se construye una matriz de rasgos donde cada fila constituye los valores que toma cada rasgo para cada territorio. A partir de esta se obtiene la matriz de similitud entre rasgos utilizando alguna medida de similitud de las reportadas en la literatura. Esta matriz será la entrada en la próxima etapa del método.

En la etapa construcción y ponderado del grafo, se construye un grafo completo donde cada rasgo constituye un nodo del grafo. Este grafo es de la clase no dirigido y ponderado y la matriz de peso es la matriz de similitud obtenida en la etapa anterior.

En la etapa de construcción del ICLA inicialmente se crea un ICLA teniendo como espacio celular el grafo de similitud obtenido en la etapa anterior. Durante la construcción del ICLA se adiciona un LA a cada nodo del grafo.

Algoritmo 1 Construir ICLA

Entrada: Un grafo de similitud $G = (V, E)$

Salida: Un ICLA $icla$

1: $icla = ICLA(G)$
2: **for all** $v_i \in V$ **do**
3: $p_i = []$
4: $action_i = accionesVecindad(v_i)$
5: **for all** $a_j \in action_i$ **do**
6: $p_i.add(pesoArista(v_i, a_j))$
7: **end for**
8: **for all** $j \in [0, len(p_i)]$ **do**
9: $p_{ij} = \frac{p_{ij}}{\sum_{p_i}}$
10: **end for**
11: $LA_{v_i} = construirLearningAutomata(action_i, p_i)$
12: $icla.addLearningAutomata(LA_{v_i})$
13: **end for**
14: **return** $icla$

Figura 2-2: *Construcción del ICLA, elaboración propia.*

El conjunto de acciones de cada LA se corresponde con los nodos que pertenecen a su vecindad. La probabilidad inicial para cada acción se obtiene a partir del peso de la arista que une a los dos nodos.

Posteriormente se obtiene una configuración inicial para el autómata y se procede con la inicialización de los parámetros para su configuración inicial.

Algoritmo 2 Inicializar ICLA

Entrada: $icla = (G, \Phi, LA, F)$

Salida: El ICLA inicializado $icla$

for all $la_i \in LA$ **do**
  if $coefIntra(la_i) < coefInter(la_i)$ **then**
    $\beta = 0$
  else
    $\beta = 1$
  end if
  $Z_i(0) = 1$
  $W_i(0) = 1 - \beta$
  $\hat{D}_i(0) = \frac{W_i(0)}{Z_i(0)}$
  $a_i = max(\hat{D}_i)$
  for all $a_j \in la_i.actions$ **do**
    $$la_i.probVector[j] = \begin{cases} la_i.probVector[j] + \alpha * (1 - la_i.probVector[j]) & \text{si } a_j = a_i \\ (1 - \alpha) * la_i.probVector[j] & \text{en otro caso} \end{cases}$$
  end for
end for
return $icla$

Figura 2-3: *Inicialización del ICLA, elaboración propia.*

Luego de inicializado el ICLA se procede a la generación de los subconjuntos de rasgos, para ello se utiliza el algoritmos ICLASC propuesto por (Pérez Betancourt, González Polanco, Febles Rodríguez, Cabrera Campos 2020).

2.4 Herramienta Informática XANGEO

Los requisitos para un sistema son las descripciones de los servicios que un sistema debe proporcionar y las restricciones sobre su funcionamiento. Estos requisitos reflejan las necesidades de los clientes de un sistema que cumple un determinado propósito (Sommerville 2015). La calidad con que se realiza la captura de los requisitos incide en todo el proceso de desarrollo del software repercutiendo en el resto de las fases de su desarrollo. Además, contribuye a tomar mejores decisiones de diseño y arquitectura.

Esta instanciación tiene como objetivo demostrar la viabilidad del método propuesto y facilitar la evaluación concreta de su idoneidad en la regionalización de territorios. El método propuesto, ha sido implementado como un complemento para el SIG QGis y cuenta con los siguientes requisitos funcionales:

2.4.1 Requisitos Funcionales

Los requisitos funcionales (RF) para un sistema describen lo que este debe hacer. Los RF son declaraciones de servicios que el sistema debe proporcionar, cómo debe reaccionar el sistema a entradas particulares y cómo debe comportarse el sistema en situaciones particulares. En algunos casos, los requisitos funcionales también pueden indicar explícitamente lo que el sistema no debe hacer (Sommerville 2015). Los requerimientos funcionales pueden ser: cálculos, detalles técnicos, manipulación de datos y otras funcionalidades específicas que se supone que un sistema debe cumplir. A continuación, se muestran los requisitos identificados:

RF 1: Importar rasgos temáticos.
RF 2: Obtener rasgos geoespaciales a través de QGis.
RF 3: Construir grafo de restricciones espaciales.
RF 4: Construir ICLA.
RF 4.1: Inicializar el ICLA.
RF 5: Generar subconjunto.
RF 6: Evaluar subconjunto.
RF 7: Construir regionalización.
RF 8: Gestionar regionalización.
RF 8.1: Eliminar regionalización.
RF 8.2: Visualizar regionalización.
RF 8.3: Adicionar regionalización.
RF 9: Exportar regionalización como imagen.
RF 10: Exportar regionalización hacia una hoja de cálculo.

2.4.2 Requisitos No Funcionales

Los requisitos no funcionales (RNF) son restricciones en los servicios o funciones que ofrece el sistema. Incluyen restricciones de tiempo, restricciones en el proceso de desarrollo y restricciones impuestas por los estándares. Los requisitos no funcionales a menudo se aplican al sistema como un todo en lugar de las características o servicios individuales del sistema (Sommerville 2015). Estas propiedades o cualidades se refieren a las características que hacen al sistema estable, usable, rápido, confiable y escalable.

A continuación, se muestran los RNF identificados:

Requisitos de Software

RNF 1: Se debe tener instalada la herramienta QGis en su versión 3.10.

RNF 2: Se debe tener instalado el GBD PostgreSQL en su versión 9.0 o superior.

RNF 3: Se debe tener instalado el módulo PostGIS en su versión 2.1.5 o superior.

Requisitos de Hardware

RNF 4: La estación de trabajo debe contar con al menos 1,0 GB de Random Access Memory (RAM, por sus siglas en inglés).

RNF 5: La capacidad mínima de espacio en disco debe ser 2.0 GB.

Requisitos de Usabilidad

RNF 6: Debe tener una interfaz gráfica visualmente atractiva para el usuario. La aplicación podrá ser usada por cualquier usuario con conocimientos básicos sobre geografía e informática. Debe mostrar mensajes al usuario que le ayuden a llevar a cabo la tarea que realiza.

2.5 Fase de Planificación

La planeación es la etapa inicial que plantea la metodología XP. En este punto se comienza a interactuar con el cliente y el resto del grupo de desarrollo para descubrir los requerimientos del sistema, se lleva a cabo el proceso de identificación y confección de las historias de usuario, así como la familiarización del equipo de trabajo con las tecnologías y herramientas seleccionadas para el desarrollo del software. El cliente establece la prioridad de cada historia

de usuario, y correspondientemente, los programadores realizan una estimación del esfuerzo necesario de cada una de ellas (Sadath, Karim, Gill 2018). Se identifican el número y tamaño de las iteraciones al igual que se plantean ajustes necesarios a la metodología según las características del proyecto. El resultado de la fase es un plan de entregas donde se realiza una estimación de las versiones que tendrá el producto en su realización, de manera tal que guíe su desarrollo (Kruchten, Fraser, Coallier 2019).

2.5.1 Historias de Usuarios

Las HU constituyen la técnica utilizada en XP para especificar los requisitos del software; en ellas el cliente describe brevemente las características que el sistema debe poseer, y se realiza una por cada característica principal del sistema. La redacción de las mismas se realiza bajo la terminología del cliente, no del desarrollador, de forma que sea clara y sencilla, sin profundizar en detalles. El tratamiento de las HU es muy dinámico y flexible, en cualquier momento pueden reemplazarse por otras más específicas o generales, añadirse nuevas o ser modificadas. Cada HU es lo suficientemente comprensible y delimitada para que los programadores puedan implementarla en unas semanas (Sadath, Karim, Gill 2018).

Luego de obtener las principales funcionalidades del sistema, se identificaron 13 HU. En la ***Tabla 2-1*** y ***Tabla 2-2*** se muestra una breve descripción de dos de ellas.

Historia de Usuario: “Importar rasgos temáticos”	
Número: 1	Nombre HU: Importar rasgos temáticos
Usuario: Experto	
Prioridad en Negocio: Alto	Riesgo en Desarrollo: Alto
Puntos Estimados: 1	Iteración Asignada: 1
Programador Responsable: Monica Frómeta Torres	
Descripción: El método debe ser capaz de importar los rasgos temáticos correspondientes a cada capa de la base cartográfica desde diferentes fuentes.	
Observaciones:	

Tabla 2-1: *Historia de Usuario: Importar rasgos temáticos.*

Historia de Usuario: “Visualizar regionalización”	
Número: 8.2	Nombre HU: Visualizar regionalización
Usuario: Experto	
Prioridad en Negocio: Alto	Riesgo en Desarrollo: Medio
Puntos Estimados: ½	Iteración Asignada: 5
Programador Responsable: Monica Frómeta Torres	
Descripción: El método debe ser capaz de mostrar el resultado del proceso de regionalización obtenido a través del subconjunto de rasgos empleado.	
Observaciones:	

Tabla 2-2: *Historia de Usuario: Visualizar Regionalización.*

2.5.2 Estimación de esfuerzos por HU

En el presente epígrafe se realiza la estimación del esfuerzo por HU, las HU deben ser programadas en un tiempo de una a tres semanas. Si la estimación es superior a tres semanas, se divide en dos o más HU. Si es menor de una semana, se combina con otra HU. Estas estimaciones permiten tener una medida de la velocidad del proyecto y ofrecen una guía a la cual ajustarse. Los resultados estimados se muestran en la ***Tabla 2-3***.

Historia de Usuario	***Puntos de Estimación (semanas)***
HU 1: Importar rasgos temáticos	1
HU 2: Obtener rasgos geoespaciales a través de QGis	1
HU 3: Construir grafo de restricciones espaciales	1
HU 4: Construir ICLA	½
HU 4.1: Inicializar el ICLA	½
HU 5: Generar subconjunto	½
HU 6: Evaluar subconjunto	½
HU 7: Construir regionalización	1
HU 8.1: Eliminar regionalización	½
HU 8.2: Visualizar regionalización	½
HU 8.3: Adicionar regionalización	1
HU 9: Exportar regionalización como imagen	1
HU 10: Exportar regionalización hacia una hoja de cálculo	1

Tabla 2-3: *Estimación de esfuerzo por HU.*

2.5.3 Plan de iteraciones

Una vez finalizadas las HU se crea el plan de iteraciones, indicando cuáles se desarrollarán en cada iteración. En la ***Tabla 2-4*** se muestra cómo quedó definido el plan de iteraciones para la solución propuesta.

Iteraciones	*Historias de Usuario a implementar*	*Duración de Iteración (semanas)*
Iteración 1	Importar rasgos temáticos Obtener rasgos geoespaciales a través de QGis	2
Iteración 2	Construir grafo de restricciones espaciales	1
Iteración 3	Construir ICLA Inicializar el ICLA	1
Iteración 4	Generar subconjunto Evaluar subconjunto	1
Iteración 5	Construir regionalización Eliminar regionalización Visualizar regionalización Adicionar regionalización	3
Iteración 6	Exportar regionalización como imagen Exportar regionalización hacia una hoja de cálculo	2
Total		**9**

Tabla 2-4: *Plan de duración de las iteraciones.*

2.5.4 Plan de entregas

El plan de entregas establece qué HU serán agrupadas para conformar una entrega, y el orden de implementación. En este plan se concentran las funcionalidades referentes a un mismo tema en módulos, esto permite un mayor entendimiento en la fase de implementación. Tiene como objetivo definir el número de liberaciones que se realizarán en el transcurso del proyecto y las iteraciones que se requieren para desarrollar cada una. De esta forma se puede trazar el plan de entrega en función de estos dos parámetros: el tiempo de desarrollo ideal y el grado de importancia para el cliente. En la ***Tabla 2-5*** se presenta el plan de entregas de la aplicación informática propuesta.

	Final de la 1ra iteración	***Final de la 2da iteración***	***Final de la 3ra iteración***	***Final de la 4ta iteración***	***Final de la 5ta iteración***	***Final de la 6ta iteración***
Módulos	2da semana de marzo	3ra semana de marzo	4ta semana de marzo	1ra semana de abril	4ta semana de abril	2da semana de mayo
Método no supervisado de selección de rasgos	v1.0	v1.1	v1.2	v1.3	v1.4	Finalizado

Tabla 2-5: *Plan de duración de las entregas.*

2.6 Fase de Diseño

La metodología de desarrollo XP plantea prácticas especializadas que accionan directamente en la realización del diseño para lograr un sistema robusto y reutilizable. Se trata en todo momento de mantener la simplicidad para crear un diseño evolutivo que va mejorando incrementalmente. Crear soluciones mínimas que exploran las posibles soluciones para un problema específico, ignorando todos los otros problemas para disminuir el riesgo.

2.6.1 Arquitectura de Software

El diseño arquitectónico se preocupa por comprender cómo debe organizarse un sistema de software y diseñar la estructura general de ese sistema. En el modelo del proceso de desarrollo de software el diseño arquitectónico es la primera etapa del proceso de diseño de software. Es el vínculo crítico entre el diseño y la ingeniería de requisitos, ya que identifica los principales componentes estructurales de un sistema y las relaciones entre ellos. El resultado del proceso de diseño arquitectónico es un modelo arquitectónico que describe cómo se organiza el sistema como un conjunto de componentes comunicantes (Sommerville 2015).

El diseño del sistema y la organización está regido por un estilo arquitectónico de Arquitectura en capas. Este enfoque en capas soporta el desarrollo incremental de sistemas, lo cual puede ser observado en la ***Figura 2-4***.

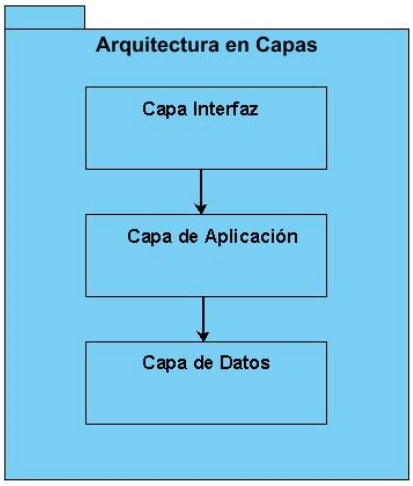


Figura 2-4: *Diagrama Arquitectura en capas, elaboración propia.*

La aplicación de este estilo arquitectónico posibilita que las funcionalidades del sistema estén organizadas en capas separadas y cada una se apoya sólo en las facilidades y los servicios ofrecidos por la capa inmediatamente debajo de ella (Sommerville 2015). La capa proporciona servicios a la capa superior, por lo que las capas de nivel más bajo representan servicios centrales que probablemente se utilizarán en todo el sistema. La capa inicial es responsable de implementar la interfaz de usuario, la segunda capa implementa la funcionalidad del sistema, la capa de base de datos, ofrece administración de transacciones y almacenamiento constante de datos.

Como patrón de arquitectura de software se escoge Modelo-Vista-Controlador, logrando que el sistema quede organizado y así tener un orden lógico en la programación del mismo.

Modelo Vista Controlador:

El patrón Modelo Vista Controlador (Model View Controller, MVC por sus siglas en inglés) es un patrón de arquitectura de software que separa los datos de una aplicación, la interfaz de usuario, y la lógica de control en tres componentes distintos. El patrón MVC se ve frecuentemente en aplicaciones web, donde la vista es la interfaz de usuario y el código es el que provee de datos dinámicos a la página; el modelo es el Sistema de Gestión de Base de Datos y la Lógica de negocio; y el controlador es el responsable de recibir los eventos de entrada desde la vista (Pressman 2013; Sommerville 2015).

En esta investigación se presenta el sistema (Sistema para el análisis espacial en salud XANGEO) que está compuesto por los módulos:

- Común provee útiles para facilita la conexión con el SIG.
- Los módulos Estratificador, Servicio y Regionalizador en los cuales se evidencia el patrón arquitectónico MVC como se muestra en la ***Figura 2-5***.

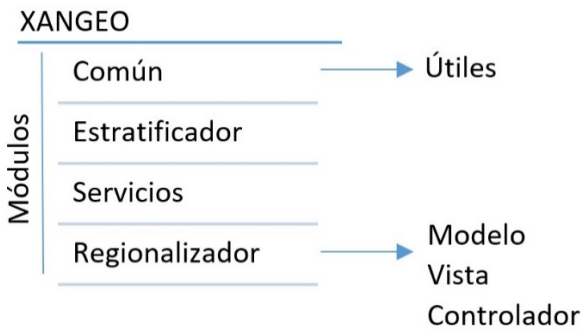


Figura 2-5: *Evidencia de la arquitectura del sistema, elaboración propia.*

Modelo: Esta es la representación específica de la información con la cual el sistema opera. Maneja los datos y controla sus transformaciones. Este no tiene conocimiento específico de los controladores y las vistas, ni siquiera contiene referencias a ellos. Es el propio sistema el que tiene encomendada la responsabilidad de mantener enlaces entre el modelo y sus vistas y notificar a las vistas cuando cambia el modelo.

Vista: Maneja la presentación visual de los datos representados por el modelo, el cual es presentado en un formato adecuado para interactuar, usualmente la interfaz de usuario.

Controlador: Responde a eventos, usualmente acciones del usuario, e invoca cambios en el modelo y probablemente en la vista.

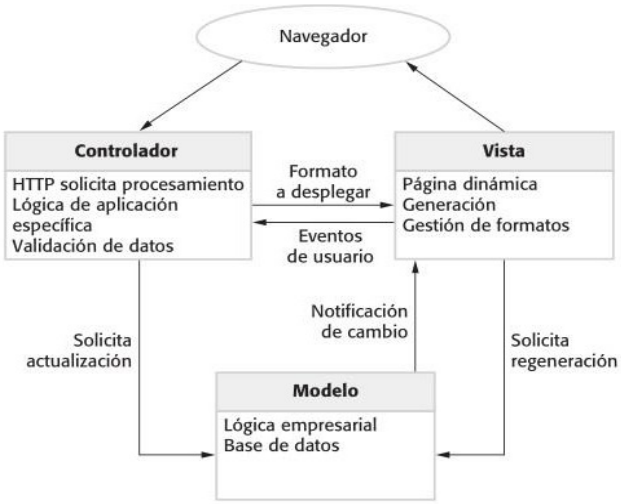


Figura 2-6: *Patrón MVC, tomado de (Sommerville 2015).*

2.6.2 Tarjetas Clase-Responsabilidad-Colaboración

La técnica Clase-Responsabilidad-Colaboración (CRC) propone una forma de trabajo, preferentemente grupal, para encontrar los objetos del dominio de la aplicación, sus responsabilidades y cómo colaboran con otros para realizar tareas. Esta técnica utiliza las llamadas tarjetas CRC. Las tarjetas CRC son una técnica de diseño orientado a objetos. Estas identifican las clases y asociaciones que participan en el diseño del sistema, las responsabilidades que debe cumplir cada clase y el establecen cómo una clase colabora con otras clases para cumplir con sus responsabilidades. Las tarjetas permiten a todos los miembros del proyecto contribuir con ideas y recopilar las mejores dentro del diseño.

En las Tabla 2-6 y Tabla 2-7 se muestran las tarjetas CRC correspondientes a las clases.

Clase: Estrato	
Responsabilidad	**Colaboración**
Crear instancias de la clase Estrato	Estratificación

Tabla 2-6: *Tarjeta CRC para la clase Estrato.*

Clase: Estratificación	
Responsabilidad	**Colaboración**
Crear instancias de la clase Estratificación	ControladoraEstratificación

Tabla 2-7: *Tarjeta CRC para la clase Estratificación.*

2.6.3 Diagrama de Clases del diseño

El diagrama de clases del diseño describe gráficamente las especificaciones de las clases de software y de las interfaces en una aplicación. Un diagrama de este tipo presenta las clases del sistema con sus relaciones estructurales y de herencia. En la Figura 2-7 se muestra el diagrama de clases de la aplicación informática propuesta.

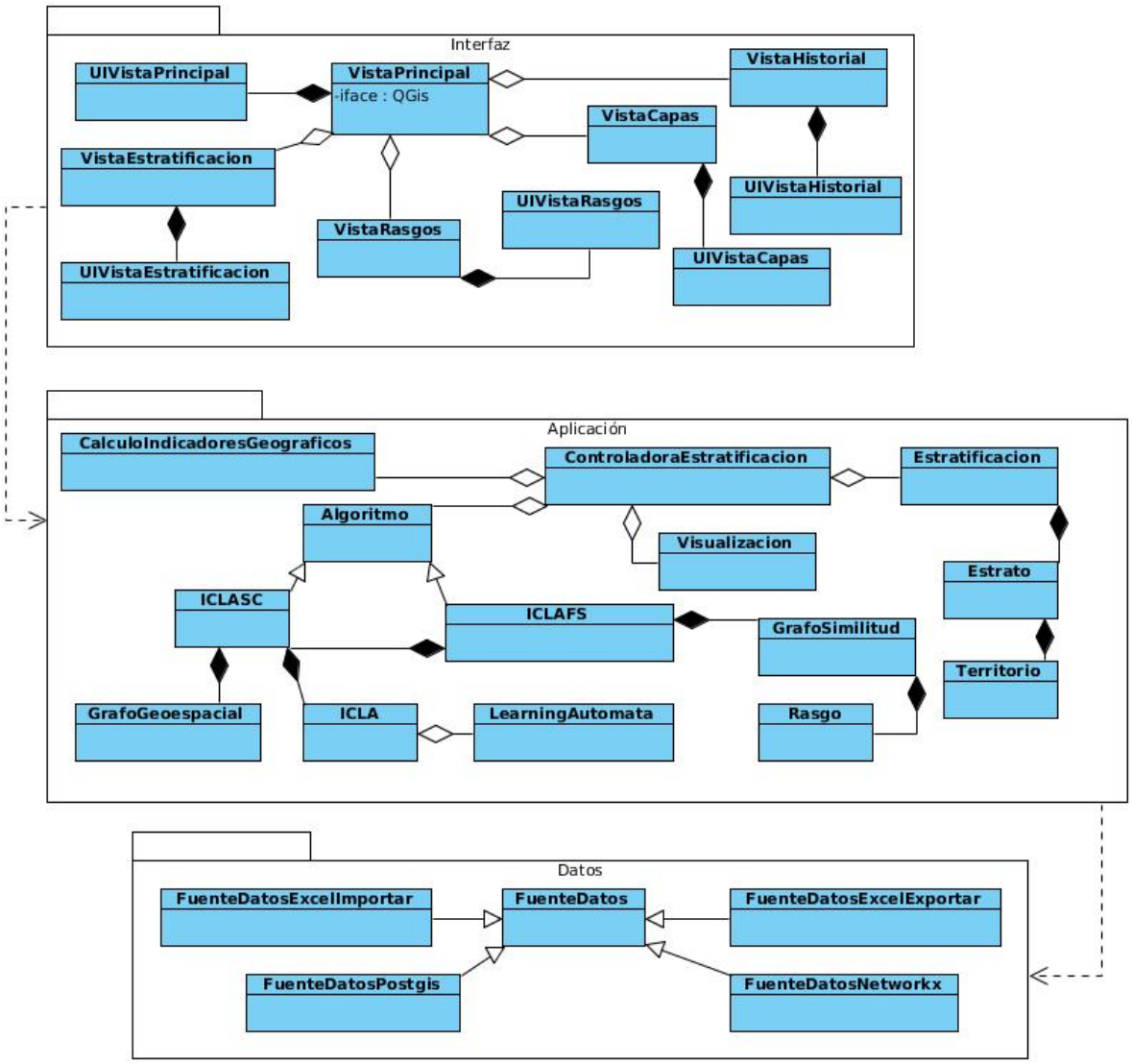


Figura 2-7: *Diagrama de clases del diseño, elaboración propia.*

2.6.4 Patrones de Diseño

Un patrón de diseño se caracteriza como “una regla de tres partes que expresa una relación entre cierto contexto, un problema y una solución” (Alexander 1979). Para el diseño de software, el contexto permite al lector entender el ambiente en el que reside el problema y qué solución sería apropiada en dicho ambiente. Un conjunto de requerimientos, incluidas limitaciones y restricciones, actúan como sistema de fuerzas que influyen en la manera en la que puede interpretarse el problema en este contexto y en cómo podría aplicarse con eficacia la solución (Pressman 2013).

Los patrones de diseño contribuyen a reutilizar diseño gráfico, identificando aspectos claves de la estructura de un diseño que puede ser aplicado en una gran cantidad de situaciones. Estos proporcionan una estructura conocida por todos los programadores, de manera que la forma de trabajar no resulte distinta entre los mismos.

2.6.4.1 Patrones Generales de Software para la Asignación

Los Patrones Generales de Software para la Asignación de Responsabilidades (GRASP, por sus siglas en inglés) son utilizados para describir los principios fundamentales del diseño y la asignación de responsabilidades UML y patrones. Entre los que se utilizaron en la solución figuran los siguientes: Experto, Creador, Controlador, Bajo acoplamiento y Alta cohesión.

Experto: este patrón plantea que se debe asignar una responsabilidad al experto en información, es decir, a la clase que tiene la información necesaria para realizar la responsabilidad. Dicho patrón se evidencia en la aplicación informática en la clase Territorio, como esta posee toda la información necesaria para calcular el aporte de riesgo de cada territorio se le asigna dicha responsabilidad. En la Figura 2-8 se muestra una imagen de dicha clase.

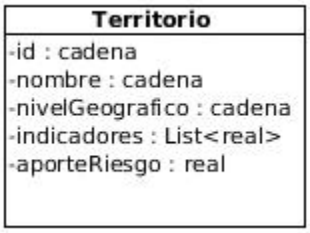


Figura 2-8: *Evidencia del patrón experto.*

Creador: define asignar a la clase B la responsabilidad de crear una instancia de clase A si se cumplen uno o más de los siguientes casos:

- B agrega objetos de A.
- B contiene objetos de A.
- B registra instancias de objetos de A.
- B utiliza estrechamente objetos de A.
- B tiene los datos de inicialización que se pasaran a un objeto de A cuando sea creado.

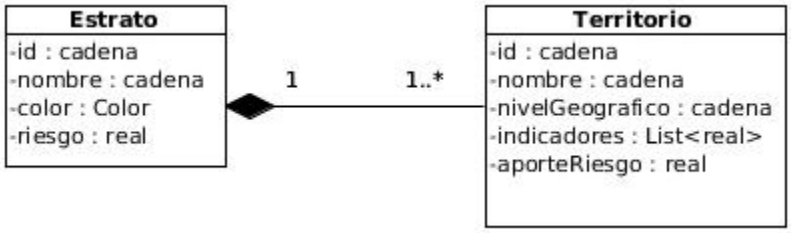


Figura 2-9: *Evidencia del patrón creador.*

Controlador: permite asignar la responsabilidad de controlar el flujo de eventos del sistema a clases específicas, facilitando la centralización de actividades. El controlador no realiza estas actividades, las delega en otras clases con las que mantiene un modelo de alta cohesión. Este patrón se evidencia en la aplicación informática en la clase ControladoraEstratificacion, a esta se le asignó la responsabilidad de manejar los eventos del sistema generados por el usuario. En la Figura 2-10 se muestra una imagen de dicha clase.

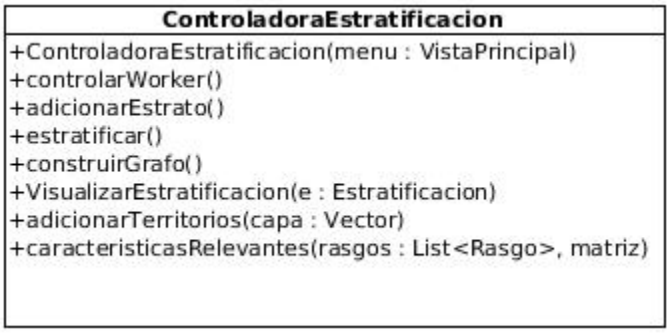


Figura 2-10: *Evidencia del patrón controlador.*

Bajo acoplamiento: asignar responsabilidades de manera que la dependencia entre clases permanezca baja. Este patrón se garantiza en la aplicación basándose en la propia arquitectura del sistema, lo que permite que las dependencias entre las clases sean muy pocas, ya que solamente las clases de una capa se pueden comunicar con las de la capa inmediatamente inferior.

Alta cohesión: La cohesión es una medida de cuán relacionadas y enfocadas están las responsabilidades de una clase. Una alta cohesión caracteriza a las clases con responsabilidades estrechamente relacionadas que no realizan un trabajo enorme. Una clase

con baja cohesión hace muchas cosas no afines o realiza trabajo excesivo. En resumen, este patrón se observa cuando una clase tiene la responsabilidad de realizar una labor dentro del sistema, no desempeñada por el resto de los componentes del diseño. Este patrón se evidencia en la aplicación informática en conjunto con el patrón bajo acoplamiento, de forma tal que cada clase realice sus acciones y se evita que otra clase realice acciones correspondientes a la clase con la que está relacionada.

2.6.4.2 Patrones del Grupo de Cuatro

Los Patrones del Grupo de Cuatro (GoF, por sus siglas en inglés) resuelven problemas específicos de diseño de software (Pressman 2013). Estos patrones se agrupan en las siguientes categorías:

Creacionales: se centran en la creación, composición y representación de objetos. Estos encierran conocimiento acerca de cuáles son las clases concretas que usa el sistema, pero al mismo tiempo ocultan la forma en la que las instancias de dichas clases se crean y agrupan. Los patrones creacionales ofrecen mecanismos que hacen más fácil la formación de las instancias de los objetos dentro de un sistema y establecen restricciones en el tipo y número de objetos que es posible crear dentro de un sistema.

Estructurales: se enfocan en problemas y soluciones asociados con la manera en la que se organizan e integran las clases y objetos para construir una estructura más grande, ayudando a establecer relaciones entre entidades dentro de un sistema.

Conductuales: Los patrones conductuales se enfocan a problemas asociados con la asignación de responsabilidades entre los objetos y a la manera en la que se efectúa la comunicación entre ellos.

Método plantilla: es un patrón de comportamiento que define en una operación el esqueleto de un algoritmo, delegando en las subclases algunos de sus pasos, esto permite que las sub-clases redefinan ciertos pasos de un algoritmo sin cambiar estructura. Este patrón se

evidencia en las clases ICLASC y ICLAFS, que heredan todas las funcionalidades de la clase Algoritmo, y redefinen los métodos en función de sus características. En la Figura 2-11 se muestra cómo se evidencia el patrón plantilla en la aplicación informática propuesta.

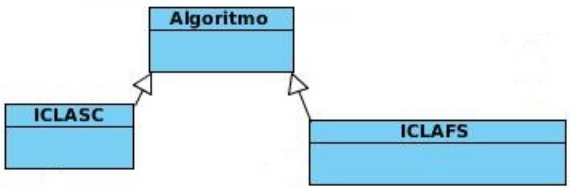


Figura 2-11: *Evidencia del patrón plantilla.*

2.7 Fase de Implementación

Luego de haber definido los elementos necesarios en la etapa de planificación y diseño se pasa a la de codificación o implementación de la aplicación, donde se da cumplimento al plan de iteraciones. En esta fase se realiza la implementación de las HU que fueron seleccionadas por cada iteración, además se crean las tareas de ingeniería para ayudar a organizar la implementación exitosa de las HU. Con el objetivo de traducir el modelo del diseño a software operativo (Pressman 2013; Kruchten, Fraser, Coallier 2019) se presenta el Diagrama de componentes el cual muestra la composición de la regionalización contenida en el complemento XANGEO.

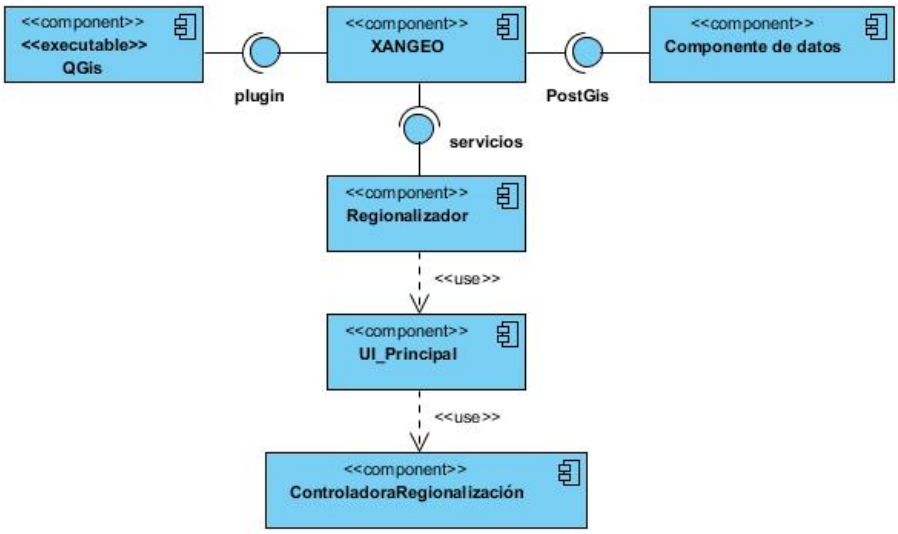


Figura 2-12: *Diagrama de componente, elaboración propia.*

2.7.1 Tareas de ingeniería

Las Historias de Usuario están compuesta por una o varias tareas de ingeniería, con el objetivo de especificar las tareas llevadas a cabo por el programador responsable de ella o ellas.

HU	**Tareas de Ingeniaría por HU**
Importar rasgos temáticos	Extraer los datos de la fuente de datos (tabla, base de datos, hoja de cálculo). Mostrar los datos en la vista estratificación.

Tabla 2-8: *Distribución de tareas de ingeniería por HU, elaboración propia.*

Tarea de Ingeniería	
Número Tarea:	Número Historia de Usuario: HU#1
Nombre: Extraer los datos de la fuente de datos (tabla, base de datos, hoja de cálculo).	
Tipo Tarea: Desarrollo	Puntos Estimados: 1
Fecha Inicio: 28/4/2020	Fecha Fin: 2/5/2020
Programador Responsable: Monica Frómeta Torres	
Descripción: Esta tarea permite extraer los datos, ya sea de una tabla, una base de datos o una hoja de cálculo.	

Tabla 2-9: *Tarea de Ingeniería Extraer los datos de la fuente de datos.*

2.7.2 Estándares de codificación

Entre los elementos más importantes que destaca XP se encuentran los estándares de codificación. XP resalta que la comunicación de los programadores es a través del código, por lo cual se hace necesario que se sigan ciertos estándares de programación para lograr un entendimiento entre dichos programadores, de forma tal, que cualquier persona del equipo de desarrollo pueda modificar el código. Además, se hace preciso que el código sea entendible para que posteriormente otros programadores puedan apoyarse en ese trabajo y desarrollen otras soluciones (Sadath, Karim, Gill 2018).

En el caso de la herramienta que se desarrolla, el estándar que se utiliza es:

Máxima longitud de las líneas.

- Todas las líneas se limitan a un máximo de 79 caracteres.

Importaciones

- Las importaciones se encuentran en líneas separadas.

Comentarios

- Se utilizan comentarios de una línea para hacer más entendible el código.
- Comentarios de una línea: comentario pequeño que solo abarca una línea y describe el código que le sigue.

Estilo de los nombres

- Clases e Interfaces: los nombres de las clases presentan la primera letra en mayúscula, en caso de ser un nombre compuesto, la inicial de cada palabra se representa en mayúscula. Se utilizan nombres simples y de alguna manera que describan el contenido, se usan palabras completas, a no ser que la abreviatura sea muy conocida.
- Métodos y variables: los nombres de cada método se representan en minúscula, en caso de ser un nombre compuesto, la inicial de la primera palabra se simboliza en minúscula, y la de las otras palabras que lo componen en mayúscula. Los nombres de las variables son cortos, pero con significados lógicos, capaces de permitir a un observador identificar su función.

2.8 Conclusiones del Capítulo

Durante este capítulo se presentó un método no supervisado para la selección de rasgos en problemas de regionalización basado en SIG y autómatas celulares que combina algoritmos y descripciones textuales identificados en enfoques precedentes con el objetivo de obtener subconjuntos óptimos de rasgos que ofrezcan mejores resultados en el proceso de regionalización. A partir de la ejecución de las etapas previstas según el paradigma empleado se arrojan las siguientes conclusiones.

La identificación de los requisitos permitió un mayor entendimiento de las necesidades del cliente. Mediante la descripción de las HU divididas por iteraciones y la planificación del esfuerzo dedicado al desarrollo en cada una de ellas, se logró una mejor organización del trabajo y el establecimiento de fechas para la culminación por cada una de las iteraciones. Las tareas de ingeniería correspondiente a cada HU permitieron la organización del trabajo

en una secuencia lógica de pasos y el estándar de codificación utilizado proporcionó un buen entendimiento y una mejor estructuración del código. La integración de técnicas de selección de rasgos en el proceso de regionalización de territorios facilita la incorporación del espacio en estudios salubristas y constituye una alternativa de análisis alineada al principio de la primera ley de la geografía. La instanciación del método como un componente para el Sistema de Información Geográfica QGIS facilita la evaluación de la viabilidad de la propuesta y lo dota de flexibilidad para integrar datos de variada naturaleza en estos estudios.

CAPÍTULO 3. VALIDACIÓN DE LA PROPUESTA

En este capítulo se presentan los principales elementos relacionados con la validación del método no supervisado para la selección de rasgos en problemas de regionalización basado en Sistemas de Información Geográfica y Teoría de Autómatas Celulares. También se discuten los resultados obtenidos a partir de la aplicación en casos de estudios, con el objetivo de comprobar que se obtienen mejores resultados a través del método propuesto. Se realizan las pruebas definidas por la metodología seleccionada, así como las pruebas unitarias para verificar el código, y las pruebas de aceptación para comprobar si al final de cada iteración se consiguió la funcionalidad requerida.

3.1 Fase de Pruebas

El ciclo de vida de desarrollo de software describe las fases para el desarrollo de software, desde su fase inicial hasta su implementación y mantenimiento. De forma similar, el proceso de pruebas de software describe la manera de cómo va a ser probado el sistema software. Llevar a cabo un adecuado proceso de pruebas de software permite identificar y corregir los defectos encontrados a lo largo del proceso de desarrollo de software, garantizando un alto grado la calidad del producto (Chillán Zulca, Lozano Pushug 2017). XP divide las pruebas en dos grupos: pruebas de aceptación, o pruebas funcionales diseñadas por el cliente final, destinadas a evaluar si al final de una iteración se consiguió la funcionalidad requerida y pruebas unitarias, encargadas de verificar el código, diseñadas por los programadores (Sharma, Hasteer 2016; Sohaib, Solanki, Dhaliwa, Hussain, Asif 2019).

3.1.1 Pruebas de Aceptación

Las pruebas de aceptación constituyen las listas de verificación que evidencian el cumplimiento de los requerimientos del sistema desde la perspectiva del usuario, y se centran en las características y funcionalidades generales del sistema, que son visibles y revisables por parte del usuario (Sadath, Karim, Gill 2018; Kruchten, Fraser, Coallier 2019). Estas pruebas se derivan de las HU que se han implementado como parte de la liberación del software. Los clientes son responsables de verificar que los resultados de estas pruebas sean correctos. Así mismo, en caso de que fallen varias pruebas, deben indicar el orden de

prioridad de resolución. Una HU no se puede considerar terminada hasta tanto pase correctamente todas las pruebas de aceptación.

Para validar que el resultado obtenido por el sistema coincide con el resultado esperado por el cliente se diseñaron un total de 13 casos de prueba de aceptación en conjunto cliente-desarrolladores. De este total, 4 arrojaron el resultado esperado mientras que 9 pruebas resultaron fallidas, las funcionalidades que respondían a estas pruebas fueron tratadas en la siguiente iteración y al volver a aplicar las pruebas de funcionalidad mostraron un resultado de 8 exitosas y 5 fallidas. En la siguiente iteración se realizó el tratamiento correspondiente a estas funcionalidades obteniéndose 10 pruebas exitosas y 3 fallidas. En la última iteración al aplicar las pruebas se obtuvieron un total 13 pruebas satisfactorias para 13 casos de prueba aplicados, como se muestra en la Figura 3-1.

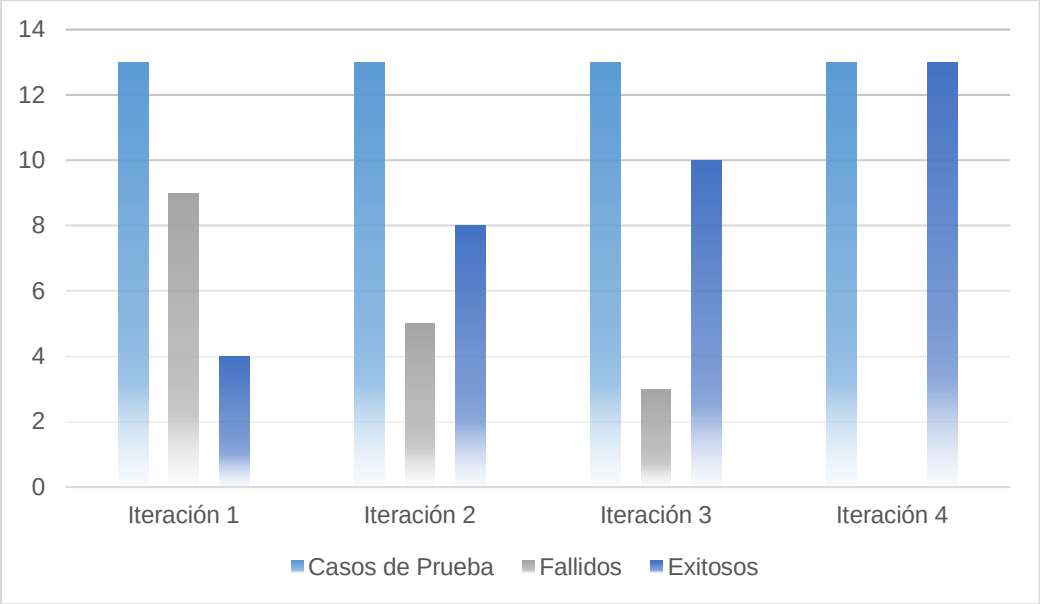


Figura 3-1: *Resultados de las pruebas de aceptación, elaboración propia.*

3.1.2 Estratificación de territorios según la diez principales causas de muerte en el año 2016

Para valorar los resultados de la solución propuesta se decide aplicar otro caso de estudio, en correspondencia con el trabajo realizado por (Pérez Betancourt, González Polanco, Febles

Rodríguez, Cabrera Campos 2018), en el cual se utiliza la división política-administrativa del año 2011, por ello se seleccionan 15 provincias y el municipio especial Isla de la Juventud.

En este caso de estudio la estratificación de territorios se enfoca como un problema de regionalización a partir de adicionar restricciones geoespaciales. Desde este enfoque, la estratificación de territorios garantiza la contigüidad espacial de los polígonos que representan a los territorios dentro de un estrato o región.

Dado un conjunto de territorios $T = \{t_1 , t_2 , \ldots , t_n \}$ la estratificación de territorios como un problema de regionalización debe determinar el conjunto de regiones R que cumplen:

i. los territorios son agregados en un número de regiones a partir de la optimización de un criterio particular de agregación

ii. dentro de una región los territorios cumplen con la restricción geoespacial definida

iii. la cantidad de regiones es menor o igual a la cantidad de territorios, $|R| \leq n$

iv. un territorio solo es asignado a una única región

Para su ejecución, se obtiene una capa vectorial desde la IDERC (Infraestructura de Datos Espaciales de la República de Cuba) con los polígonos que representan a cada territorio escogido para el análisis. Parten de la hipótesis de relación de las enfermedades con el espacio y seleccionan como variables las diez principales causas de muerte de Cuba en el año 2016.

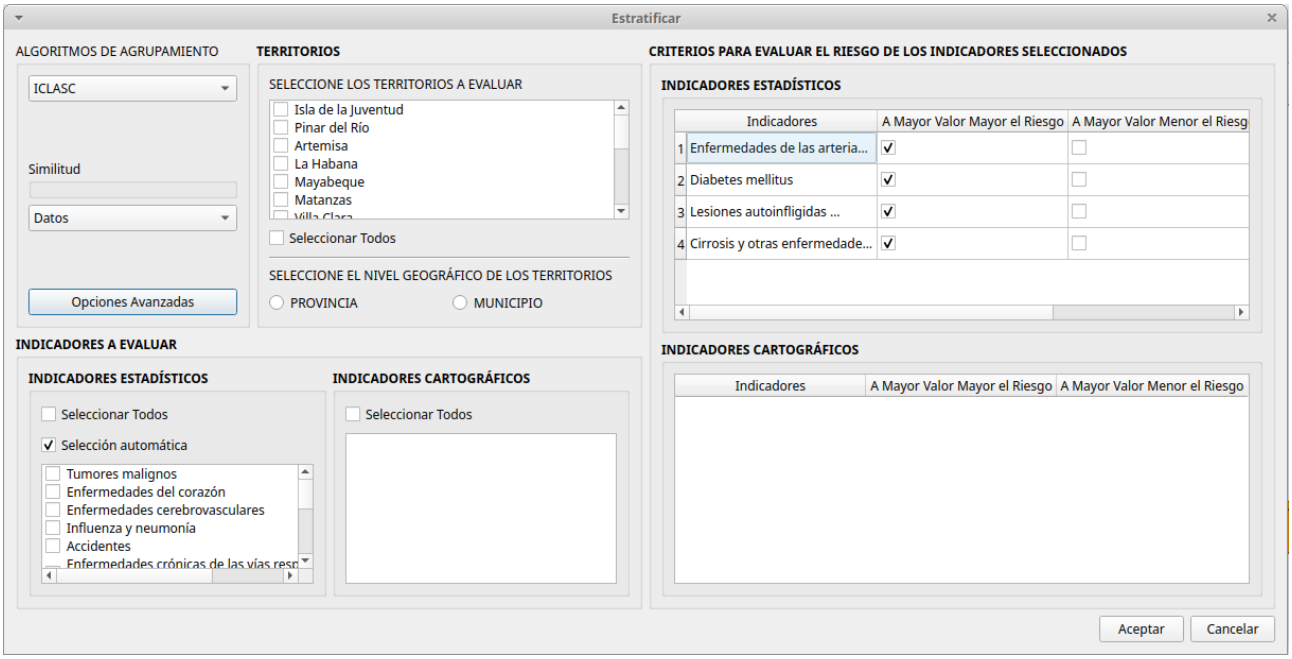
Los indicadores de estas variables por territorios se obtienen del Anuario estadístico de salud del mencionado año. Para realizar el estudio descrito utilizando el método propuesto en la presente investigación, se emplea el algoritmo de agrupamiento ICLASC propuesto por (Pérez Betancourt, González Polanco, Febles Rodríguez, Cabrera Campos 2020). Luego se realiza una estratificación cada subconjunto de rasgos que se obtiene por el método propuesto. Se realiza además una evaluación del comportamiento del agrupamiento utilizando índices de validación externos en comparación con el resultado obtenido en (Pérez Betancourt, González Polanco, Febles Rodríguez, Cabrera Campos 2020).

Para determinar la similitud entre los rasgos se utilizó la función *chi2_kernel* y como resultado se obtuvo tres subconjuntos de rasgos como se muestra en la **¡Error! No se encuentra el origen de la referencia.**.

Subconjunto 1	Subconjunto 2	Subconjunto 3
Tumores malignos Enfermedades del corazón	Enfermedades cerebro-vasculares Influenza y neumonía Accidentes Enfermedades crónicas de las vías respiratorias inferiores	Enfermedades de las arterias, arteriolas y vasos capilares Diabetes mellitus Lesiones auto-infligidas intencionalmente Cirrosis y otras enfermedades crónicas del hígado

Tabla 3-1: *Subconjuntos obtenidos.*

3.1.3 Resultados de la estratificación de territorios según las principales causas de muerte en el 2016



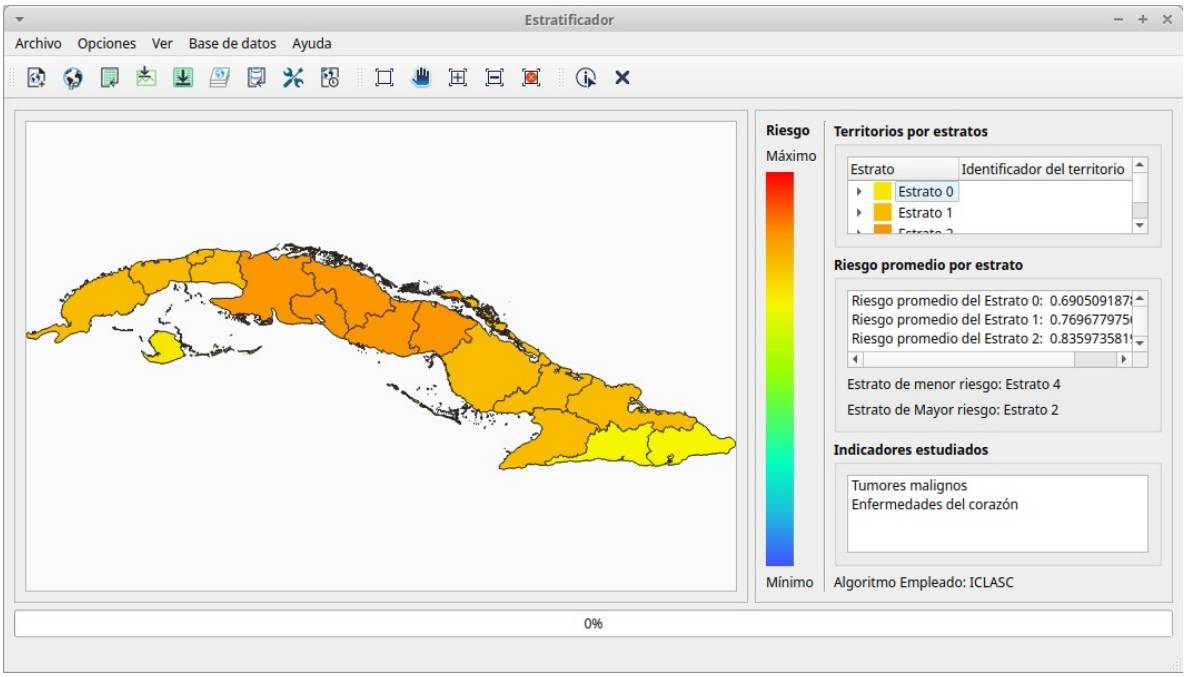


Figura 3-2: *Mapa temático estratificado con ICLASC y el subconjunto 1.*

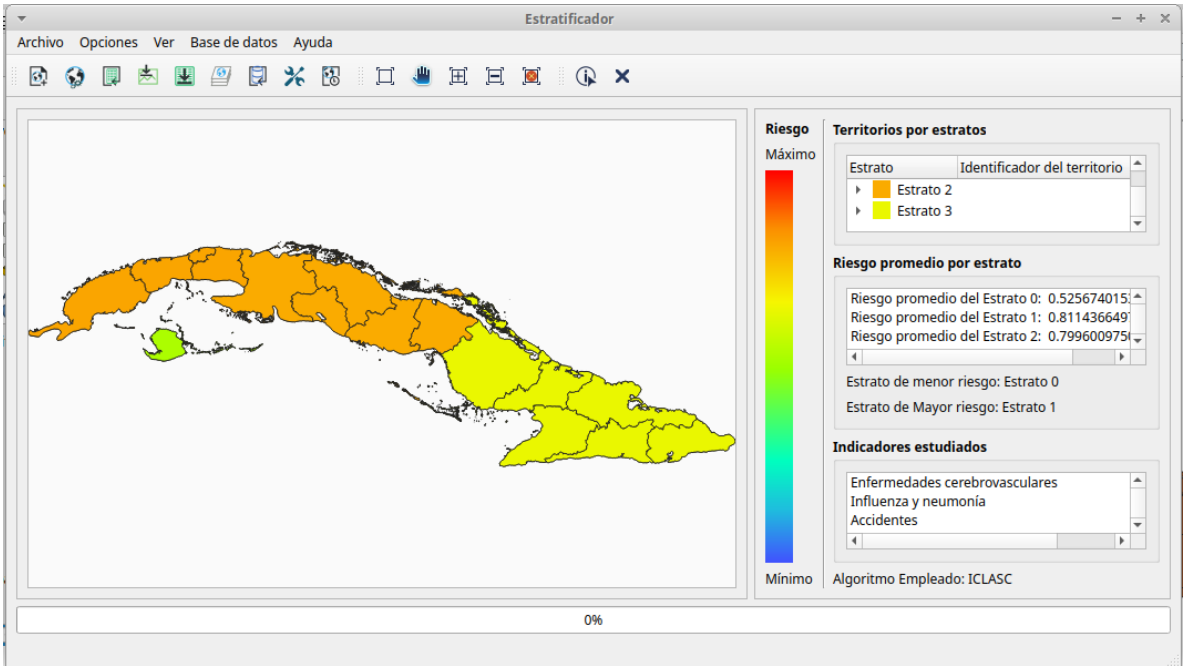


Figura 3-3: *Mapa temático estratificado con ICLASC y el subconjunto 2.*

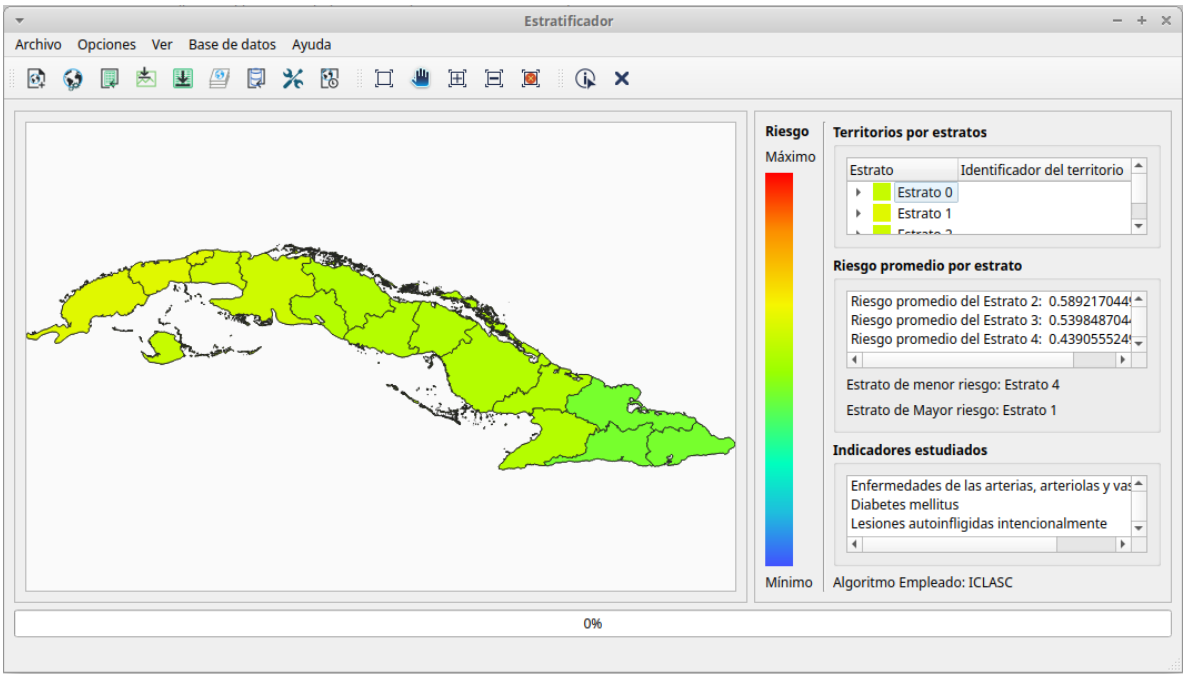


Figura 3-4: *Mapa temático estratificado con ICLASC y el subconjunto 3.*

Los resultados de los índices de validación internos para los estudios realizados se muestran en la Tabla 3-2. Primeramente, se observa que con las funciones de distancia geométricas se obtienen siempre grupos más compactos que solo con la utilización de la componente temática. Para el resto de los índices los resultados son competitivos en comparación con la función temática y destaca el desempeño de la conectividad y la distancia en el espacio.

Estratificación	calinski_harabaz	silhouette_score
ICLASC+sub 1	2,100	0,003
ICLASC+sub 2	2,600	0,070
ICLASC+sub 3	1,500	-0,035

Tabla 3-2: *Resultado de evaluar índices de validación internos, elaboración propia.*

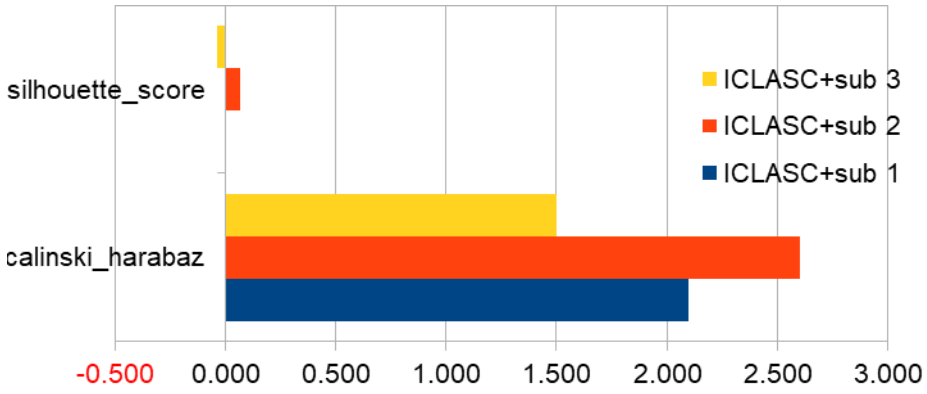


Tabla 3-3: *Evaluación de índices de validación internos, elaboración propia.*

Posteriormente se calculan los índices de validación tomando como referencia el estudio presentado por (Pérez Betancourt, González Polanco, Febles Rodríguez, Cabrera Campos 2020), evidenciando un buen resultado en cuanto a la precisión por lo que las funciones de distancia geométricas para este estudio obtienen grupos más compactos sin afectar la precisión.

Estratificación	precision	jaccard	fowlkes_mallows	v_measure	completeness	homogeneity
ICLASC+ subconjunto 1	1,00	0,88	0,86	0,91	0,84	1,00
ICLASC+ subconjunto 2	1,00	1,00	1,00	1,00	1,00	1,00
ICLASC+ subconjunto 3	0,65	0,56	0,48	0,59	0,54	0,64

Tabla 3-4: *Resultado de evaluar índices de validación externos, elaboración propia.*

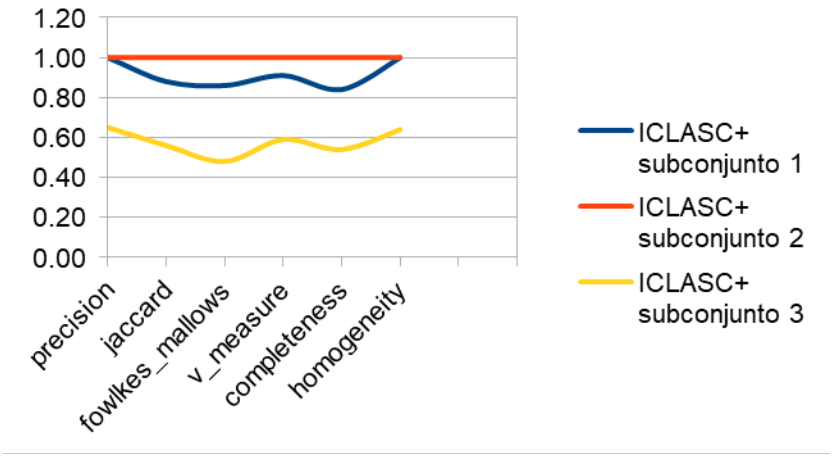


Tabla 3-5: *Evaluación de índices de validación externos, elaboración propia.*

3.1.4 Análisis de los resultados

En la literatura consultada recomiendan emplear pruebas no paramétricas (no presuponen una distribución de probabilidad para los datos) basadas en rangos de Friedman seguida de comparaciones múltiples (Benavoli, Corani, Mangili, Zaffalon 2015) para identificar pares de algoritmos que difieran significativamente. Se recomiendan cuando los tamaños de muestra son pequeños, especialmente cuando el número de instancias es menor que 30 (Rousseau 2016). En estos casos se emplea como parámetro de centralización la mediana, que es aquel punto para el que el valor de X está el 50 por ciento de las veces por debajo y el 50 por ciento por encima. Las ventajas del enfoque no paramétrico son: no promedia las medidas tomadas en diferentes conjuntos de datos, no asume normalidad de los medios muestrales y es robusto a los valores atípicos.

Friedman es una prueba de significación de hipótesis nula, por lo tanto, controla el error de Tipo I, es decir, la probabilidad de rechazar la hipótesis nula cuando es verdadera. Se le considera como un análisis de varianza no paramétrico para un diseño experimental en bloques. Por lo tanto hay que cumplir con dos suposiciones:

- se tiene k muestras relacionadas
- la escala de medición de la variable a probar está al menos en escala ordinal

Wilcoxon es considerada como una alternativa a la prueba de t (t de Student) para dos muestras pareadas. El procedimiento de ambas pruebas se basa en el cálculo de diferencias Di=xi–yi entre pares de observaciones, pero en la prueba de Wilcoxon se asignan rangos a las

diferencias. En esta prueba la hipótesis se plantea en torno a la mediana de las diferencias (Md), mientras que en la prueba de t se plantea sobre la media de diferencias (_D), en la literatura consultada se enuncia una eficiencia de esta prueba del 95 por ciento para muestras pequeñas. Se utilizan las pruebas no paramétricas porque a pesar de basarse en determinadas suposiciones, no parten de la base de que los datos analizados adoptan una distribución normal, el tamaño de la muestra es pequeño. De las pruebas disponibles en la literatura se aplica Friedman, que es una prueba para comprobar la igualdad de tratamientos en medidas repetidas, para contrastar la hipótesis nula de igualdad entre esos tratamientos.

La prueba estadística Friedman obtiene un estadístico de prueba es igual a 0.80, por lo que se puede rechazar la hipótesis nula para niveles de significación superiores a 0.67. Al cinco por ciento de nivel de significación se rechaza la hipótesis de que existen diferencias significativas y se concluye que todos los subconjuntos tienen un comportamiento semejante en la estratificación. Se realizan comparaciones dos a dos empleando la prueba de Wilcoxon de rangos con signo. Wilcoxon permite determinar si la diferencia entre la magnitud de las diferencias positivas entre los valores de las dos variables y la magnitud de las diferencias negativas es estadísticamente significativa. Los resultados obtenidos y que se muestran en la Tabla 3-6 no se evidencian diferencias significativas para ninguna.

	ICLASC+sub 1 - ICLASC+sub 2	ICLASC+sub 1- ICLASC+sub 3	ICLASC+sub 2- ICLASC+sub 3
Z	10.0	15.0	21.0
Sig. Asint.	0.26	0.67	0.15

Tabla 3-6: *Resultados de la prueba de Wilcoxon, elaboración propia.*

3.2 Conclusiones del Capítulo

En el presente capítulo con la aplicación de las pruebas de aceptación se pudo detectar, documentar y corregir las no conformidades existentes en el sistema implementado. La realización de estas pruebas permitió verificar el correcto funcionamiento del sistema y el cumplimiento de los requisitos del cliente. En el capítulo se muestran los resultados obtenidos al aplicar el método propuesto en un caso de estudio, Estratificación de territorios sobre las principales causas de muerte en Cuba durante el año 2016. Se llegó a la conclusión de que es

posible incorporar medidas de distancia geométricas y mantener un desempeño competitivo en relación con otros trabajos publicados. Al evaluar los índices de validación tanto internos como externos se pudo comprobar que si se incorporan medidas de distancia geométricas se obtienen grupos más compactos sin afectar la precisión de la clasificación. El análisis exploratorio de datos espaciales arrojó evidencias que permiten considerar un mejor desempeño en clasificadores con los criterios propuestos para identificar estratos con dependencia espacial y por tanto más compactos.

CONCLUSIONES

Como resultados de la presente investigación se obtuvo un método no supervisado para la selección de rasgos en problemas de regionalización utilizando SIG que contribuye a la detección de fenómenos locales y globales. En función de los resultados obtenidos se arribó a las siguientes conclusiones:

a) A partir de la sistematización de los principales referentes teóricos que sustentan la presente investigación, se confirma que las propuestas para la regionalización reportadas en la literatura presentan limitaciones para la incorporación de la componente espacial en el proceso de regionalización de territorios.
b) La definición del marco teórico referencial de la investigación relacionado con la regionalización y técnicas de selección de rasgos, fundamentaron la necesidad de desarrollar un plugin que se adapte a los objetivos expuestos y satisfaga las necesidades del país.
c) El estudio de numerosas literaturas, en conjunto con las necesidades de la investigación trajo consigo la utilización de la teoría de autómatas celulares para el desarrollo de la propuesta.
d) La identificación de los constructos dentro de la minería de datos geoespaciales facilitó la descripción adecuada del problema y su solución a partir del método propuesto, que además integra los enfoques aportados en investigaciones precedentes con relación a estudios sobre regionalización y la componente espacial de los datos en el espacio de solución del problema.
e) La integración de la solución propuesta al sistema Qgis facilitó la realización de la regionalización y análisis de la distribución espacial de los fenómenos.
f) Las pruebas aplicadas para la verificación de la solución informática y la valoración de los resultados a través de un caso de estudio demostraron que el sistema cumple con los requisitos definidos, garantizando su correcto funcionamiento.

REFERENCIAS BIBLIOGRÁFICAS

ABOUSAEIDI, Mohammad, FAUZI, Rosmadi and MUHAMAD, Rusnah, 2016. Geographic Information System (GIS) modeling approach to determine the fastest delivery routes. *Saudi Journal of Biological Sciences*. 1 September 2016. Vol. 23, no. 5, p. 555–564. DOI 10.1016/j.sjbs.2015.06.004.

ABUALIGAH, Laith Mohammad, KHADER, Ahamad Tajudin and HANANDEH, Essam Said, 2018. A new feature selection method to improve the document clustering using particle swarm optimization algorithm. *Journal of Computational Science*. 1 March 2018. Vol. 25, p. 456–466. DOI 10.1016/j.jocs.2017.07.018.

ADAMS, Matthew D., KANAROGLOU, Pavlos S. and COULIBALY, Paulin, 2016. Spatially constrained clustering of ecological units to facilitate the design of integrated water monitoring networks in the St. Lawrence Basin. *International Journal of Geographical Information Science*. 1 February 2016. Vol. 30, no. 2, p. 390–404. DOI 10.1080/13658816.2015.1089442.

AHMADI, Ali, RAMAZANI, Rashid, REZAGHOLI, Tahereh and YAVARI, Parvin, 2018. Incidence pattern and spatial analysis of breast cancer in Iranian women: Geographical Information System applications. *Eastern Mediterranean Health Journal*. 1 April 2018. Vol. 24, no. 4, p. 360–367. DOI 10.26719/2018.24.4.360.

AKSAC, Alper, ÖZYER, Tansel and ALHAJJ, Reda, 2019. CutESC: Cutting edge spatial clustering technique based on proximity graphs. *Pattern Recognition*. 1 December 2019. Vol. 96, p. 106948. DOI 10.1016/j.patcog.2019.06.014.

ALENE, Kefyalew Addis and CLEMENTS, Archie C. A., 2019. Spatial clustering of notified tuberculosis in Ethiopia: A nationwide study. *PLOS ONE*. 9 August 2019. Vol. 14, no. 8, p. e0221027. DOI 10.1371/journal.pone.0221027.

ALEXANDER, Christopher, 1979. *The Timeless Way of Building*. Oxford University Press.

ISBN 978-0-19-502402-9.
Google-Books-ID: H6CE9hlbO8sC

ANGUIX, Alvaro and DIAZ, Laura, 2008. gvSIG: A GIS desktop solution for an open SDI. *Journal of Geography and Regional Planning* [online]. 2008. [Accessed 24 February 2020]. Available from: https://www.researchgate.net/publication/228968906_gvSIG_A_GIS_desktop_solution_for_an_open_SDI

ARNIELLA PÉREZ, Angela María, 2008. *Utilización del sistema de información geográfica para determinar el comportamiento territorial de los factores de riesgo que influyen en la morbilidad por hepatitis viral A en la cabecera municipal de Güines.* Universidad de La Habana.

ATLURI, Gowtham, KARPATNE, Anuj and KUMAR, Vipin, 2018. *Spatio-Temporal Data Mining: A Survey of Problems and Methods* [online]. 22 August 2018. Association for Computing Machinery. [Accessed 8 July 2020]. Available from: https://doi.org/10.1145/3161602

ATURINDE, Augustus, FARNAGHI, Mahdi, PILESJÖ, Petter and MANSOURIAN, Ali, 2019. Spatial analysis of HIV-TB co-clustering in Uganda. *BMC Infectious Diseases*. 12 July 2019. Vol. 19, no. 1, p. 612. DOI 10.1186/s12879-019-4246-2.

AYUDIANI, Venny Larasati and AKBAR, Saiful, 2017. An extensible tool for spatial clustering. In: [online]. IEEE. November 2017. p. 1–6. ISBN 978-1-5386-1449-5. Available from: http://ieeexplore.ieee.org/document/8285848/

B, Sana, SIDDIQUI, Isma Farrah and ARAIN, Qasim Ali, 2019. Analyzing Students' Academic Performance through Educational Data Mining. *3C Tecnología. Glosas de innovación aplicadas a la pyme.* 17 May 2019. P. 402–421.

BAÇÃO, Fernando, LOBO, Victor and PAINHO, Marco, 2005. Self-organizing Maps as Substitutes for K-Means Clustering. In: *Computational Science – ICCS 2005*. Berlin, Heidelberg: Springer. 2005. p. 476–483. Lecture Notes in Computer Science. ISBN 978-3-

540-32118-7.

BEAUCHEMIN, Mario, 2019. Semi-supervised map regionalization for categorical data. *International Journal of Remote Sensing*. 17 December 2019. Vol. 40, no. 24, p. 9401–9411. DOI 10.1080/2150704X.2019.1633485.

BEIRANVAND, Reza, KARIMI, Asrin, DELPISHEHD, Ali, SAYEHMIRI, Kourosh, SOLEIMANI, Samira and GHALAVANDI, Shahnaz, 2016. Correlation Assessment of Climate and Geographic Distribution of Tuberculosis Using Geographical Information System (GIS). *Iranian Journal of Public Health*. January 2016. Vol. 45, no. 1, p. 86–93.

BELTRÁN, Pedro, 2018. Qué es una herramienta CASE. [online]. 2018. [Accessed 20 February 2020]. Available from: https://www.academia.edu/28037284/Qu%C3%A9_es_una_herramienta_CASE

BENAVOLI, Alessio, CORANI, Giorgio, MANGILI, Francesca and ZAFFALON, Marco, 2015. A Bayesian nonparametric procedure for comparing algorithms. In: *International Conference on Machine Learning* [online]. PMLR. 1 June 2015. p. 1264–1272. [Accessed 6 September 2020]. Available from: http://proceedings.mlr.press/v37/benavoli15.html

BI, Qifang, AZMAN, Andrew S., SATTER, Syed Moinuddin, KHAN, Azharul Islam, AHMED, Dilruba, RIAJ, Altaf Ahmed, GURLEY, Emily S. and LESSLER, Justin, 2016. Micro-scale Spatial Clustering of Cholera Risk Factors in Urban Bangladesh. *PLOS Neglected Tropical Diseases*. 11 February 2016. Vol. 10, no. 2, p. e0004400. DOI 10.1371/journal.pntd.0004400.

BIANCHI, Gianpiero, BRUNI, Renato, REALE, Alessandra and SFORZI, Fabio, 2016. A min-cut approach to functional regionalization, with a case study of the Italian local labour market areas. *Optimization Letters*. June 2016. Vol. 10, no. 5, p. 955–973. DOI 10.1007/s11590-015-0980-6.

BRANTLEY, Mary D., DAVIS, Nicole L., GOODMAN, David A., CALLAGHAN, William M. and BARFIELD, Wanda D., 2017. Perinatal regionalization: a geospatial view of

perinatal critical care, United States, 2010–2013. *American Journal of Obstetrics and Gynecology*. 1 February 2017. Vol. 216, no. 2, p. 185.e1-185.e10. DOI 10.1016/j.ajog.2016.10.011.

BROWDY, Michelle H., 1990. Simulated Annealing: An Improved Computer Model for Political Redistricting. *Yale Law & Policy Review*. 1990. Vol. 8, no. 1, p. 163–179. JSTOR

BYFUGLIEN, J. and NORDGÅRD, A., 1973. Region-Building — a Comparison of Methods. *Norsk Geografisk Tidsskrift - Norwegian Journal of Geography*. 1 January 1973. Vol. 27, no. 2, p. 127–151. DOI 10.1080/00291957308621875.

CAMPBELL, Jonathan and SHIN, Michael, 2018. Geographic Information System Basics. [online]. 26 February 2018. [Accessed 8 July 2020]. Available from: https://openlibrary-repo.ecampusontario.ca/jspui/handle/123456789/441

CAMPOS, Saúl González and MARTÍNEZ, Luis Felipe Fernández, 2015. Programación Extrema: Prácticas, Aceptación y Controversia. *Cultura Científica y Tecnológica* [online]. 16 June 2015. Vol. 0, no. 15. [Accessed 20 February 2020]. Available from: http://erevistas.uacj.mx/ojs/index.php/culcyt/article/view/512

CANO ROJAS, Alberto and ROJAS MATAS, Angela, 2016. Autómatas celulares y aplicaciones. *Revista Iberoamericana de Educación Matemática* [online]. 2016. [Accessed 19 February 2020]. Available from: https://www.researchgate.net/publication/304582327_Automatas_celulares_y_aplicaciones

CÁRCELES-ÁLVAREZ, Alberto, ORTEGA-GARCÍA, Juan A., LÓPEZ-HERNÁNDEZ, Fernando A., OROZCO-LLAMAS, Mayra, ESPINOSA-LÓPEZ, Blanca, TOBARRA-SÁNCHEZ, Esther and ALVAREZ, Lizbeth, 2017. Spatial clustering of childhood leukaemia with the integration of the Paediatric Environmental History. *Environmental Research*. 1 July 2017. Vol. 156, p. 605–612. DOI 10.1016/j.envres.2017.04.019.

CHANDRASHEKAR, Girish and SAHIN, Ferat, 2014. A survey on feature selection methods. *Computers and Electrical Engineering*. 1 January 2014. Vol. 40, no. 1, p. 16–28.

DOI 10.1016/j.compeleceng.2013.11.024.

CHHABRA, Karan R. and DIMICK, Justin B., 2016. Strategies for Improving Surgical Care: When Is Regionalization the Right Choice? *JAMA Surgery*. 1 November 2016. Vol. 151, no. 11, p. 1001–1002. DOI 10.1001/jamasurg.2016.1059.

CHILLÁN ZULCA, Mirian Soraya and LOZANO PUSHUG, Nelly Marisol, 2017. Elaboración de una guía de procedimientos para la fase de pruebas en el desarrollo de software. [online]. 7 September 2017. [Accessed 25 August 2020]. Available from: http://bibdigital.epn.edu.ec/handle/15000/18767
Accepted: 2017-09-07T20:27:11Z

CORTI, Paolo, KRAFT, Thomas J., MATHER, Stephen Vincent and PARK, Bborie, 2014. *PostGIS Cookbook*. Packt Publishing Ltd. ISBN 978-1-84951-867-3. Google-Books-ID: zCaxAgAAQBAJ

CUÉLLAR LUNA, Liliam and GUTIÉRREZ SOTO, Tania, 2014. Desarrollo de la geografía médica o de la salud en Cuba. *Revista Cubana de Higiene y Epidemiología*. December 2014. Vol. 52, no. 3, p. 388–401.

DELGADO ACOSTA, Hilda, GONZÁLEZ MORENO, Lídice, VALDÉS GÓMEZ, María, HERNÁNDEZ MALPICA, Sara, MONTENEGRO CALDERÓN, Tamara and RODRÍGUEZ BUERGO, Delfín, 2015. Estratificación de riesgo de tuberculosis pulmonar en consejos populares del municipio Cienfuegos. *MediSur*. April 2015. Vol. 13, no. 2, p. 275–284.

DRESCH, Aline, LACERDA, Daniel Pacheco and ANTUNES, José Antônio Valle, 2015. Design Science Research. In: *Design Science Research: A Method for Science and Technology Advancement* [online]. Cham: Springer International Publishing. p. 67–102. [Accessed 25 February 2020]. ISBN 978-3-319-07374-3. Available from: https://doi.org/10.1007/978-3-319-07374-3_4

DUEÑAS FERNÁNDEZ, Raúl, 2016. Regionalization of health services for medical care: an example from the Cardiocentro Ernesto Che Guevara. *CorSalud (Revista de Enfermedades Cardiovasculares)*. 19 December 2016. Vol. 8, no. 4, p. 248–256.

DUQUE, Juan Carlos, ARTÍS, Manuel and RAMOS, Raúl, 2006. The ecological fallacy in a time series context: evidence from Spanish regional unemployment rates. *Journal of Geographical Systems*. 1 October 2006. Vol. 8, no. 4, p. 391–410. DOI 10.1007/s10109-006-0033-x.

DUQUE, Juan Carlos and CHURCH, R. L., 2004. A new heuristic model for designing analytical regions. In: *North American Meeting of the Regional Science Association International*. Seattle. 2004.

DUQUE, Juan Carlos, 2004. *Design of homogenous territorial units. A methodological proposal and applications* [online]. Universitat de Barcelona. [Accessed 25 February 2020]. ISBN 978-84-693-8466-4. Available from: http://diposit.ub.edu/dspace/handle/2445/35354

ESNAASHARI, M. and MEYBODI, M. R., 2008. A Cellular Learning Automata Based Clustering Algorithm for Wireless Sensor Networks. [online]. October 2008. [Accessed 24 February 2020]. DOI info:doi/10.1166/sl.2008.m146. Available from: https://www.ingentaconnect.com/content/asp/senlet/2008/00000006/00000005/art00009

ESNAASHARI, Mehdi and MEYBODI, M. R., 2010. Data aggregation in sensor networks using learning automata. *Wireless Networks*. 1 April 2010. Vol. 16, no. 3, p. 687–699. DOI 10.1007/s11276-009-0162-5.

ESNAASHARI, Mehdi and MEYBODI, Mohammad Reza, 2018. Dynamic irregular cellular learning automata. *Journal of Computational Science*. 1 January 2018. Vol. 24, p. 358–370. DOI 10.1016/j.jocs.2017.08.012.

FERNÀNDEZ FRAGA, Santiago, 2014. Autómatas Celulares y su Aplicación en Computación. In: *ResearchGate* [online]. 2014. [Accessed 19 February 2020]. Available from: https://www.researchgate.net/publication/267511537_Automatas_Celulares_y_su_Aplicacion_en_Computacion

FISCHER, Kurt W., 1980. A theory of cognitive development: The control and construction

of hierarchies of skills. *Psychological Review*. 1980. Vol. 87, no. 6, p. 477–531. DOI 10.1037/0033-295X.87.6.477.

FLETCHER, Stephanie and CAPRARELLI, Graziella, 2016. Application of GIS technology in public health: successes and challenges. *Parasitology*. 2 February 2016. Vol. 1, p. 1–15. DOI 10.1017/S0031182015001869.

FRANÇA, Urbano L. and MCMANUS, Michael L., 2018. Trends in Regionalization of Hospital Care for Common Pediatric Conditions. *Pediatrics* [online]. 1 January 2018. Vol. 141, no. 1. [Accessed 8 July 2020]. DOI 10.1542/peds.2017-1940. Available from: https://pediatrics.aappublications.org/content/141/1/e20171940

GARFINKEL, R. S. and NEMHAUSER, G. L., 1970. Optimal Political Districting by Implicit Enumeration Techniques. *Management Science*. 1 April 1970. Vol. 16, no. 8, p. B-495. DOI 10.1287/mnsc.16.8.B495.

GÉ VAILLANT, Raniel, 2020. *Sistema para la georreferenciación y análisis de la distribución espacial de los tumores malignos en Cuba*.

GEORGE, John, LAMAR, Bruce and WALLACE, Chris, 1997. Political District Determination Using Large-Scale Network Optimization. *Socio-Economic Planning Sciences*. 1 March 1997. Vol. 31, p. 11–28. DOI 10.1016/S0038-0121(96)00016-X.

GHAVIPOUR, Mina and MEYBODI, Mohammad Reza, 2017. Irregular cellular learning automata-based algorithm for sampling social networks. *Engineering Applications of Artificial Intelligence*. 1 March 2017. Vol. 59, p. 244–259. DOI 10.1016/j.engappai.2017.01.004.

GÓMEZ, Alveiro Rosado, DUARTE, Alexander Quintero and GÜEVARA, Cesar Daniel Meneses, 2014. Desarrollo ágil de software aplicando programación extrema. *Revista Ingenio Universidad Francisco de Paula Santander Ocaña*. 10 February 2014. Vol. 5, no. 1, p. 24–29.

GONZÁLEZ POLANCO, Liset, 2019. *Método de estratificación de territorios basado en Sistemas de Información Geográfica y medidas de similitud geométrica*. Universidad de las Ciencias Informaticas.

HATCHUEL, Armand, LE MASSON, Pascal, REICH, Yoram and SUBRAHMANIAN, Eswaran, 2018. Design theory: a foundation of a new paradigm for design science and engineering. *Research in Engineering Design*. 1 January 2018. Vol. 29, no. 1, p. 5–21. DOI 10.1007/s00163-017-0275-2.

HE, Weixiong, LING, Haifeng, ZHANG, Zhanliang and GONG, Congcong, 2018. Multi-objective spatially constrained clustering for regionalization with particle swarm optimization. *International Journal of Geographical Information Science*. 3 April 2018. Vol. 32, no. 4, p. 827–846. DOI 10.1080/13658816.2017.1418363.

HE, Xiyan, BEAUSEROY, Pierre and SMOLARZ, André, 2015. Dynamic Feature Subspaces Selection for Decision in a Nonstationary Environment. *International Journal of Pattern Recognition and Artificial Intelligence*. 29 April 2015. Vol. 29, no. 06, p. 1551009. DOI 10.1142/S021800141551009X.

HELBIG, Robert E., ORR, Patrick K. and ROEDIGER, Robert R., 1972. Political redistricting by computer. *Communications of the ACM*. 1 August 1972. Vol. 15, no. 8, p. 735–741. DOI 10.1145/361532.361543.

HEŘMANOVSKÝ, M., HAVLÍČEK, V., HANEL, M. and PECH, P., 2017. Regionalization of runoff models derived by genetic programming. *Journal of Hydrology*. April 2017. Vol. 547, p. 544–556. DOI 10.1016/j.jhydrol.2017.02.018.

HERRERA, Francisco, 2006. Técnicas de reducción de datos en KDD. El uso de Algoritmos Evolutivos para la Selección de Instancias. *Actas del I Seminario Sobre Sistemas Inteligentes (SSI'06), Universidad Rey Juan Carlos, Madrid* [online]. 2006. [Accessed 19 February 2020]. Available from: https://www.academia.edu/2932699/T%C3%A9cnicas_de_reducci%C3%B3n_de_datos_en_KDD._El_uso_de_Algoritmos_Evolutivos_para_la_Selecci%C3%B3n_de_Instancias

HÖWER, Daniel, OBERST, Christian A. and MADLENER, Reinhard, 2019. General regionalization heuristic to map spatial heterogeneity of macroeconomic impacts: The case of the green energy transition in NRW. *Utilities Policy*. June 2019. Vol. 58, p. 166–174. DOI 10.1016/j.jup.2019.05.002.

HU, Rongyao, ZHU, Xiaofeng, CHENG, Debo, HE, Wei, YAN, Yan, SONG, Jingkuan and ZHANG, Shichao, 2017. Graph self-representation method for unsupervised feature selection. *Neurocomputing*. 12 January 2017. Vol. 220, p. 130–137. DOI 10.1016/j.neucom.2016.05.081.

HUANG, Chi-Chieh, TAM, Tuen Yee Tiffany, CHERN, Yinq-Rong, LUNG, Shih-Chun Candice, CHEN, Nai-Tzu and WU, Chih-Da, 2018. Spatial Clustering of Dengue Fever Incidence and Its Association with Surrounding Greenness. *International Journal of Environmental Research and Public Health*. September 2018. Vol. 15, no. 9, p. 1869. DOI 10.3390/ijerph15091869.

HUANG, Pei, MA, Zhenjun, XIAO, Longzhu and SUN, Yongjun, 2019. Geographic Information System-assisted optimal design of renewable powered electric vehicle charging stations in high-density cities. *Applied Energy*. 1 December 2019. Vol. 255, p. 113855. DOI 10.1016/j.apenergy.2019.113855.

JOHNSTON, R. J., 1968. Choice in Classification: The Subjectivity Of objective Methods. *Annals of the Association of American Geographers*. 1968. Vol. 58, no. 3, p. 575–589. DOI 10.1111/j.1467-8306.1968.tb01653.x.

KATAYAMA, Kayoko, YOKOYAMA, Kazuhito, YAKO-SUKETOMO, Hiroko, OKAMOTO, Naoyuki, TANGO, Toshiro and INABA, Yutaka, 2014. Breast Cancer Clustering in Kanagawa, Japan: A Geographic Analysis. *Asian Pacific Journal of Cancer Prevention*. 15 January 2014. Vol. 15, no. 1, p. 455–460. DOI 10.7314/APJCP.2014.15.1.455.

KAVAKIOTIS, Ioannis, TSAVE, Olga, SALIFOGLOU, Athanasios, MAGLAVERAS, Nicos, VLAHAVAS, Ioannis and CHOUVARDA, Ioanna, 2017. Machine Learning and Data Mining Methods in Diabetes Research. *Computational and Structural Biotechnology*

Journal. 1 January 2017. Vol. 15, p. 104–116. DOI 10.1016/j.csbj.2016.12.005.

KIM, Kamyoung, DEAN, Denis J., KIM, Hyun and CHUN, Yongwan, 2016. Spatial optimization for regionalization problems with spatial interaction: a heuristic approach. *International Journal of Geographical Information Science*. 3 March 2016. Vol. 30, no. 3, p. 451–473. DOI 10.1080/13658816.2015.1031671.

KORTE, George B., 2001. *The GIS Book*. Cengage Learning. ISBN 978-0-7668-2820-9. Google-Books-ID: _C6oPvJ5S_EC

KRUCHTEN, Philippe, FRASER, Steven and COALLIER, François, 2019. *Agile Processes in Software Engineering and Extreme Programming : 20th International Conference, XP 2019, Montréal, QC, Canada, May 21–25, 2019, Proceedings* [online]. Springer Nature. [Accessed 8 July 2020]. Available from: https://library.oapen.org/handle/20.500.12657/23099

LANKFORD, Philip, 1969. Regionalization: Theory and Alternative Algorithms. *Geographical Analysis*. 1969. Vol. 1, p. 196–212. DOI 10.1111/j.1538-4632.1969.tb00615.x.

LAST, Anna, BURR, Sarah, ALEXANDER, Neal, HARDING-ESCH, Emma, ROBERTS, Chrissy H., NABICASSA, Meno, CASSAMA, Eunice Teixeira da Silva, MABEY, David, HOLLAND, Martin and BAILEY, Robin, 2017. Spatial clustering of high load ocular Chlamydia trachomatis infection in trachoma: a cross-sectional population-based study. *Pathogens and Disease* [online]. 31 July 2017. Vol. 75, no. 5. [Accessed 8 July 2020]. DOI 10.1093/femspd/ftx050. Available from: https://academic.oup.com/femspd/article/75/5/ftx050/3791466

LEBECHEREL, Laure, ANDRÉASSIAN, Vazken and PERRIN, Charles, 2016. On evaluating the robustness of spatial-proximity-based regionalization methods. *Journal of Hydrology*. August 2016. Vol. 539, p. 196–203. DOI 10.1016/j.jhydrol.2016.05.031.

LI, Jundong, GUO, Ruocheng, LIU, Chenghao and LIU, Huan, 2019. Adaptive Unsupervised Feature Selection on Attributed Networks. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* [online]. Anchorage, AK,

USA: Association for Computing Machinery. 25 July 2019. p. 92–100. [Accessed 8 July 2020]. KDD '19. ISBN 978-1-4503-6201-6. Available from: https://doi.org/10.1145/3292500.3330856

LI, Jundong and LIU, Huan, 2017. Challenges of Feature Selection for Big Data Analytics. *IEEE Intelligent Systems*. March 2017. Vol. 32, no. 2, p. 9–15. DOI 10.1109/MIS.2017.38.

LIANG, Ching-Ping, CHEN, Jui-Sheng, CHIEN, Yi-Chi and CHEN, Ching-Fang, 2018. Spatial analysis of the risk to human health from exposure to arsenic contaminated groundwater: A kriging approach. *Science of The Total Environment*. 15 June 2018. Vol. 627, p. 1048–1057. DOI 10.1016/j.scitotenv.2018.01.294.

LIU, Xin, WANG, Xiangyu, WRIGHT, Graeme, CHENG, Jack C. P., LI, Xiao and LIU, Rui, 2017. A State-of-the-Art Review on the Integration of Building Information Modeling (BIM) and Geographic Information System (GIS). *ISPRS International Journal of Geo-Information*. February 2017. Vol. 6, no. 2, p. 53. DOI 10.3390/ijgi6020053.

LU, Huijuan, CHEN, Junying, YAN, Ke, JIN, Qun, XUE, Yu and GAO, Zhigang, 2017. A hybrid feature selection algorithm for gene expression data classification. *Neurocomputing*. 20 September 2017. Vol. 256, p. 56–62. DOI 10.1016/j.neucom.2016.07.080.

LUMPKIN, Stephanie and STITZENBERG, Karyn, 2018. Regionalization and Its Alternatives. *Surgical Oncology Clinics*. 1 October 2018. Vol. 27, no. 4, p. 685–704. DOI 10.1016/j.soc.2018.05.009.

LUO, Minnan, NIE, Feiping, CHANG, Xiaojun, YANG, Yi, HAUPTMANN, Alexander G. and ZHENG, Qinghua, 2018. Adaptive Unsupervised Feature Selection With Structure Regularization. *IEEE Transactions on Neural Networks and Learning Systems*. April 2018. Vol. 29, no. 4, p. 944–956. DOI 10.1109/TNNLS.2017.2650978.

MACLEOD, Liam C., CANNON, Shannon S., KO, Oliver, SCHADE, George R., WRIGHT, Jonathan L., LIN, Daniel W., HOLT, Sarah K., GORE, John L. and DASH, Atreya, 2018. Disparities in Access and Regionalization of Care in Testicular Cancer. *Clinical Genitourinary Cancer*. 1 August 2018. Vol. 16, no. 4, p. e785–e793.

DOI 10.1016/j.clgc.2018.02.014.

MACMILLAN, W., 2001. Redistricting in a GIS environment: An optimisation algorithm using switching-points. *Journal of Geographical Systems*. 1 August 2001. Vol. 3, no. 2, p. 167–180. DOI 10.1007/PL00011473.

MACMILLAN, William and PIERCE, Todd, 1994. Optimizaction modelling in a GIS framework: the problem of political redistricting. In: *Spatial Analysis And GIS*. Taylor & Francis.

MAFARJA, Majdi, ALJARAH, Ibrahim, FARIS, Hossam, HAMMOURI, Abdelaziz I., AL-ZOUBI, Ala' M. and MIRJALILI, Seyedali, 2019. Binary grasshopper optimisation algorithm approaches for feature selection problems. *Expert Systems with Applications*. 1 March 2019. Vol. 117, p. 267–286. DOI 10.1016/j.eswa.2018.09.015.

MARAVALLE, Maurizio and SIMEONE, Bruno, 1995. A spanning tree heuristic for regional clustering. *Communications in Statistics - Theory and Methods*. 1 January 1995. Vol. 24, no. 3, p. 625–639. DOI 10.1080/03610929508831512.

MARGULES, C R, FAITH, D P and BELBIN, L, 1985. An Adjacency Constraint in Agglomerative Hierarchical Classifications of Geographic Data. *Environment and Planning A: Economy and Space*. 1 March 1985. Vol. 17, no. 3, p. 397–412. DOI 10.1068/a170397.

MASON, L. G. and GU, XueDuo, 1986. Learning Automata Models for Adaptive Flow Control in Packet-Switching Networks. In: *Adaptive and Learning Systems: Theory and Applications* [online]. Boston, MA: Springer US. p. 213–227. [Accessed 24 February 2020]. ISBN 978-1-4757-1895-9. Available from: https://doi.org/10.1007/978-1-4757-1895-9_14

MEHROTRA, Anuj, JOHNSON, Ellis L. and NEMHAUSER, George L., 1998. An Optimization Based Heuristic for Political Districting. *Management Science*. 1 August 1998. Vol. 44, no. 8, p. 1100–1114. DOI 10.1287/mnsc.44.8.1100.

MENDOZA PEÑA, Dayana, 2016. *Extensión de la herramienta Visual Paradigm for UML para la evaluación y corrección de Diagramas de Casos de Uso* [online]. Universidad de las Ciencias Informaticas. [Accessed 20 February 2020]. Available from: https://www.researchgate.net/publication/305160646_Extension_de_la_herramienta_Visual_Paradigm_for_UML_para_la_evaluacion_y_correccion_de_Diagramas_de_Casos_de_Uso

MIAO, Jianyu and NIU, Lingfeng, 2016. A Survey on Feature Selection. *Procedia Computer Science*. 1 January 2016. Vol. 91, p. 919–926. DOI 10.1016/j.procs.2016.07.111.

MIDDLETON, Richard Stephen, 2006. *Geographical distillation: Application of the p-median, traveling salesman, and regionalization problems* [online]. [Accessed 25 February 2020]. Available from: https://search.proquest.com/openview/9e48cc2e9a270b99bd4cfe0016019a80/1?pq-origsite=gscholar&cbl=18750&diss=y

MILLS, Edwin S., 1967. An Aggregative Model of Resource Allocation in a Metropolitan Area. *The American Economic Review*. 1967. Vol. 57, no. 2, p. 197–210. JSTOR

MIRANDA, Leandro, 2017. RegK-Means: A Clustering Algorithm Using Spatial Contiguity Constraints for Regionalization Problems. . 2017. P. 31–36. DOI 10.1109/BRACIS.2017.70.

MOISE, Imelda K. and RUIZ, Marilyn O., 2016. Hospitalizations for Substance Abuse Disorders Before and After Hurricane Katrina: Spatial Clustering and Area-Level Predictors, New Orleans, 2004 and 2008. *Preventing Chronic Disease*. 13 October 2016. Vol. 13, p. 160107. DOI 10.5888/pcd13.160107.

MORRONE, Juan J., 2018. The spectre of biogeographical regionalization. *Journal of Biogeography*. 2018. Vol. 45, no. 2, p. 282–288. DOI 10.1111/jbi.13135.

NAFIUL ISLAM, Quazi, 2015. *Mastering PyCharm* [online]. [Accessed 20 February 2020]. Available from: https://books.google.com/books/about/Mastering_PyCharm.html?hl=es&id=MPh_CwAAQ

BAJ

NARENDRA, Kumpati S. and THATHACHAR, Mandayam A. L., 2012. *Learning Automata: An Introduction*. Courier Corporation. ISBN 978-0-486-49877-5.

NIE, Feiping, ZHU, Wei and LI, Xuelong, 2016. Unsupervised Feature Selection with Structured Graph Optimization. In: *Thirtieth AAAI Conference on Artificial Intelligence* [online]. 21 February 2016. [Accessed 8 July 2020]. Available from: https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12180

OLAYA, Victor, 2016. *Sistemas de Información Geográfica*.

OMRANI, Hichem, TAYYEBI, Amin and PIJANOWSKI, Bryan, 2017. Integrating the multi-label land-use concept and cellular automata with the artificial neural network-based Land Transformation Model: an integrated ML-CA-LTM modeling framework. *GIScience & Remote Sensing*. 4 May 2017. Vol. 54, no. 3, p. 283–304. DOI 10.1080/15481603.2016.1265706.

OPENSHAW, S and RAO, L, 1995. Algorithms for Reengineering 1991 Census Geography. *Environment and Planning A: Economy and Space*. 1 March 1995. Vol. 27, no. 3, p. 425–446. DOI 10.1068/a270425.

OPENSHAW, S., 1973. Insoluble problems in shopping model calibration when the trip pattern is not known. *Regional Studies*. 1 December 1973. Vol. 7, no. 4, p. 367–371. DOI 10.1080/09595237300185391.

OPENSHAW, Stan, BLAKE, Marcus and WYMER, Colin, 1995. Using neurocomputing methods to classify. *Innovations in GIS*. 1995. P. 97.

ORDÓÑEZ, Mariuxi Paola Zea, RÍOS, Jimmy Rolando Molina and CASTILLO, Fausto Fabían Redrován, 2017. *ADMINISTRACIÓN DE BASES DE DATOS CON POSTGRESQL*.

3Ciencias. ISBN 978-84-946684-6-3.

PACKARD, Norman H. and WOLFRAM, Stephen, 1985. Two-dimensional cellular automata. *Journal of Statistical Physics*. 1 March 1985. Vol. 38, no. 5, p. 901–946. DOI 10.1007/BF01010423.

PALACIO, Manuel Alfredo Pech, 2002. Adaptacion y Uso de Mineria de Datos Espaciales y no Espaciales. [online]. 16 May 2002. [Accessed 19 February 2020]. Available from: http://catarina.udlap.mx/u_dl_a/tales/documentos/msp/pech_p_ma/

PEÑA SUÁREZ, Alfonso, 2017. Modelo para la caracterización del delito en la ciudad de Bogotá, aplicando técnicas de minería de datos espaciales. [online]. 16 June 2017. [Accessed 9 July 2020]. Available from: http://repository.udistrital.edu.co/handle/11349/6519

PÉREZ BETANCOURT, Yadian Guillermo, GONZÁLEZ POLANCO, Liset, FEBLES RODRÍGUEZ, Juan Pedro and CABRERA CAMPOS, Alcides, 2018. Propuestas para el análisis geoespacial en estudios salubristas. *Revista Cubana de Ciencias Informáticas*. June 2018. Vol. 12, no. 2, p. 44–57.

PÉREZ BETANCOURT, Yadian Guillermo, GONZÁLEZ POLANCO, Liset, FEBLES RODRÍGUEZ, Juan Pedro and CABRERA CAMPOS, Alcides, 2020. Cellular Automata Based Method for Territories Stratification in Geographic Information Systems. In: *Advances in Emerging Trends and Technologies*. Cham: Springer International Publishing. 2020. p. 507–517. Advances in Intelligent Systems and Computing. ISBN 978-3-030-32022-5.

PÉREZ BETANCOURT, Yadian Guillermo, GONZÁLEZ POLANCO, Liset and FEBLES RODRÍGUEZ, Juan Pedro, 2018. XANGEO: SISTEMA INFORMÁTICO PARA EL ANÁLISIS GEOESPACIAL EN ESTUDIOS SALUBRISTAS. *Informática Habana 2020* [online]. 2018. [Accessed 19 February 2020]. Available from: http://www.informaticahabana.cu/es/node/3959

PÉREZ, Luis Ismael, 2014. RELACIONES ESPACIALES BÁSICAS. [online]. 2014. [Accessed 24 February 2020]. Available from:

https://es.slideshare.net/LuisIsmaelPrez/conceptos-espaciounidad-1

PERRUCHET, Christophe, 1983. Constrained agglomerative hierarchical classification. *Pattern Recognition*. 1 January 1983. Vol. 16, no. 2, p. 213–217. DOI 10.1016/0031-3203(83)90024-9.

PRESSMAN, R.S., 2013. *Ingeniería del software UN ENFOQUE PRÁCTICO*. Octava Edición. The McGraw-Hill.

REMESEIRO, Beatriz and BOLON-CANEDO, Veronica, 2019. A review of feature selection methods in medical applications. *Computers in Biology and Medicine*. 1 September 2019. Vol. 112, p. 103375. DOI 10.1016/j.compbiomed.2019.103375.

REQUIA, Weeberb J., KOUTRAKIS, Petros, ROIG, Henrique L., ADAMS, Matthew D. and SANTOS, Cleide M., 2016. Association between vehicular emissions and cardiorespiratory disease risk in Brazil and its variation by spatial clustering of socio-economic factors. *Environmental Research*. 1 October 2016. Vol. 150, p. 452–460. DOI 10.1016/j.envres.2016.06.027.

REZAPOOR MIRSALEH, Mehdi and MEYBODI, Mohammad Reza, 2016. A new memetic algorithm based on cellular learning automata for solving the vertex coloring problem. *Memetic Computing*. 1 September 2016. Vol. 8, no. 3, p. 211–222. DOI 10.1007/s12293-016-0183-4.

REZVANIAN, Alireza and MORADABADI, Behnaz, 2019. *Learning Automata Approach for Social Networks* [online]. Springer. [Accessed 19 February 2020]. Studies in Computational Intelligence. Available from: http://ce.aut.ac.ir/~meybodi/paper/Rezvanian-Moradabadi-----Learning%20Automata%20Approach-%20for%20Social%20Networks-ToC-978-3-030-10767-3-.pdf

ROBERTO, 2016. Aplicaciones de los SIG en la salud pública. *Gis&Beers* [online]. 26 November 2016. [Accessed 24 February 2020]. Available from: http://www.gisandbeers.com/aplicaciones-de-los-sig-en-la-salud-publica/

RODRÍGUEZ, Alonso and MARÍA, Ana, 2019. Análisis filosófico-metodológico de la Investigación en Educación: la perspectiva de las Ciencias de Diseño. [online]. 2019. [Accessed 9 July 2020]. Available from: https://ruc.udc.es/dspace/handle/2183/24527

RODRÍGUEZ, Dr C. Romel Vázquez, 2018. Uso de sistemas de información geográfica libres para la protección del medio ambiente. Caso de estudio: manipulación de mapas ráster con datos climáticos. *Universidad y Sociedad*. 27 February 2018. Vol. 10, no. 2, p. 158–164.

ROUSSEAU, Judith, 2016. On the Frequentist Properties of Bayesian Nonparametric Methods. *Annual Review of Statistics and Its Application*. 2016. Vol. 3, no. 1, p. 211–231. DOI 10.1146/annurev-statistics-041715-033523.

SADATH, Lipsa, KARIM, Kayvan and GILL, Stephen, 2018. Extreme programming implementation in academia for software engineering sustainability. In: *2018 Advances in Science and Engineering Technology International Conferences (ASET)*. February 2018. p. 1–6.

SALAZAR, Jose H., GOLDSTEIN, Seth D., YANG, Jingyan, GAUSE, Colin, SWARUP, Abhishek, HSIUNG, Grace E., RANGEL, Shawn J., GOLDIN, Adam B. and ABDULLAH, Fizan, 2016. Regionalization of Pediatric Surgery: Trends Already Underway. *Annals of Surgery*. June 2016. Vol. 263, no. 6, p. 1062–1066. DOI 10.1097/SLA.0000000000001666.

SEGAL, M. and WEINBERGER, D. B., 1977. Turfing. *Operations Research*. 1 June 1977. Vol. 25, no. 3, p. 367–386. DOI 10.1287/opre.25.3.367.

SHARMA, Pooja and HASTEER, Nitasha, 2016. Analysis of linear sequential and extreme programming development methodology for a gaming application. In: *2016 International Conference on Communication and Signal Processing (ICCSP)*. April 2016. p. 1916–1920.

SHAWENO, Debebe, KARMAKAR, Malancha, ALENE, Kefyalew Addis, RAGONNET, Romain, CLEMENTS, Archie CA, TRAUER, James M., DENHOLM, Justin T. and MCBRYDE, Emma S., 2018. Methods used in the spatial analysis of tuberculosis epidemiology: a systematic review. *BMC Medicine*. 18 October 2018. Vol. 16, no. 1, p. 193.

DOI 10.1186/s12916-018-1178-4.

SHEIKHPOUR, Razieh, SARRAM, Mehdi Agha, GHARAGHANI, Sajjad and CHAHOOKI, Mohammad Ali Zare, 2017. A Survey on semi-supervised feature selection methods. *Pattern Recognition*. 1 April 2017. Vol. 64, p. 141–158. DOI 10.1016/j.patcog.2016.11.003.

SMITH, Jennifer L., AUALA, Joyce, TAMBO, Munyaradzi, HAINDONGO, Erastus, KATOKELE, Stark, UUSIKU, Petrina, GOSLING, Roly, KLEINSCHMIDT, Immo, MUMBENGEGWI, Davis and STURROCK, Hugh J. W., 2017. Spatial clustering of patent and sub-patent malaria infections in northern Namibia: Implications for surveillance and response strategies for elimination. *PLOS ONE*. 18 August 2017. Vol. 12, no. 8, p. e0180845. DOI 10.1371/journal.pone.0180845.

SOHAIB, Osama, SOLANKI, Hiralkumari, DHALIWA, Navkiran, HUSSAIN, Walayat and ASIF, Muhammad, 2019. Integrating design thinking into extreme programming. *Journal of Ambient Intelligence and Humanized Computing*. 1 June 2019. Vol. 10, no. 6, p. 2485–2492. DOI 10.1007/s12652-018-0932-y.

SOLORIO FERNÁNDEZ, Saúl, CARRASCO OCHOA, J. Ariel and MARTÍNEZ TRINIDAD, José Fco., 2020. A review of unsupervised feature selection methods. *Artificial Intelligence Review*. February 2020. Vol. 53, no. 2, p. 907–948. DOI 10.1007/s10462-019-09682-y.

SOLORIO-FERNÁNDEZ, Saúl, CARRASCO-OCHOA, J. Ariel and MARTÍNEZ-TRINIDAD, José Fco., 2018. Ranking Based Unsupervised Feature Selection Methods: An Empirical Comparative Study in High Dimensional Datasets. In: *Advances in Soft Computing* [online]. Cham: Springer International Publishing. p. 205–218. ISBN 978-3-030-04490-9. Available from: http://link.springer.com/10.1007/978-3-030-04491-6_16

SOMMERVILLE, Ian, 2015. *Software Engineering*. 10th Edition. Pearson Education. ISBN 978-1-292-09613-1.

SPENCE, N. A., 1968. A multifactor uniform regionalization of British counties on the basis

of employment data for 1961. *Regional Studies* [online]. 1968. [Accessed 25 February 2020]. DOI 10.1080/09595236800185071. Available from: https://rsa.tandfonline.com/doi/abs/10.1080/09595236800185071
world

SURYANTARA, I. Gusti Ngurah and ANDRY, Johanes Fernandes, 2018. Development of Medical Record With Extreme Programming SDLC. *IJNMT (International Journal of New Media Technology)*. 30 June 2018. Vol. 5, no. 1, p. 47–53. DOI 10.31937/ijnmt.v5i1.706.

TENBENSEL, 2016. Health System Regionalization - the New Zealand Experience. *Healthcarepapers*. 1 January 2016. Vol. 16, no. 1, p. 27–33. DOI 10.12927/hcpap.2016.24771.

TOBLER, Waldo, 1979. Smooth Pycnophylactic Interpolation for Geographic Regions. *Journal of the American Statistical Association*. 1 February 1979. Vol. 74, p. 519–30. DOI 10.1080/01621459.1979.10481647.

Unified Modeling Language, 2019. [online]. [Accessed 20 February 2020]. Available from: https://www.uml.org/what-is-uml.htm

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO, 2015. Las Relaciones Espaciales - Las Relaciones Espaciales.pdf. [online]. 2015. [Accessed 19 February 2020]. Available from: http://lae.ciga.unam.mx/arcgis/M1/Las%20Relaciones%20Espaciales.pdf

VAFASHOAR, Reza and MEYBODI, Mohammad Reza, 2019. Reinforcement learning in learning automata and cellular learning automata via multiple reinforcement signals. *Knowledge-Based Systems*. April 2019. Vol. 169, p. 1–27. DOI 10.1016/j.knosys.2019.01.021.

VALI, Y., RASHIDIAN, A., JALILI, M., OMIDVARI, A. H. and JEDDIAN, A., 2017. Effectiveness of regionalization of trauma care services: a systematic review. *Public Health*. 1 May 2017. Vol. 146, p. 92–107. DOI 10.1016/j.puhe.2016.12.006.

VENABLE, John, PRIES-HEJE, Jan and BASKERVILLE, Richard, 2016. FEDS: a Framework for Evaluation in Design Science Research. *European Journal of Information Systems*. 1 January 2016. Vol. 25, no. 1, p. 77–89. DOI 10.1057/ejis.2014.36.

VIOLINI, María Lucía, 2014. *Selección de características* [online]. Tesis. Universidad Nacional de La Plata. [Accessed 19 February 2020]. Available from: http://sedici.unlp.edu.ar/handle/10915/63236

WANG, Fahui, 2020. Why public health needs GIS: a methodological overview. *Annals of GIS*. 2 January 2020. Vol. 26, no. 1, p. 1–12. DOI 10.1080/19475683.2019.1702099.

WEBSTER, R. and BURROUGH, P. A., 1972. Computer-Based Soil Mapping of Small Areas from Sample Data. *Journal of Soil Science*. 1972. Vol. 23, no. 2, p. 210–221. DOI 10.1111/j.1365-2389.1972.tb01654.x.

WELKE, Karl F., PASQUALI, Sara K., LIN, Paul, BACKER, Carl L., OVERMAN, David M., ROMANO, Jennifer C. and KARAMLOU, Tara, 2020. Regionalization of Congenital Heart Surgery in the United States. *Seminars in Thoracic and Cardiovascular Surgery*. 1 March 2020. Vol. 32, no. 1, p. 128–137. DOI 10.1053/j.semtcvs.2019.09.005.

WU, Huanyu, WANG, Jiayuan, DUAN, Huabo, OUYANG, Lei, HUANG, Wenke and ZUO, Jian, 2016. An innovative approach to managing demolition waste via GIS (geographic information system): a case study in Shenzhen city, China. *Journal of Cleaner Production*. 20 January 2016. Vol. 112, p. 494–503. DOI 10.1016/j.jclepro.2015.08.096.

YAMAOKA, Kazue, SUZUKI, Masako, INOUE, Mariko, ISHIKAWA, Hirono and TANGO, Toshiro, 2020. Spatial clustering of suicide mortality and associated community characteristics in Kanagawa prefecture, Japan, 2011–2017. *BMC Psychiatry*. 18 February 2020. Vol. 20, no. 1, p. 74. DOI 10.1186/s12888-020-2479-7.

YUAN, Shuai, TAN, Pang-Ning, CHERUVELIL, Kendra Spence, COLLINS, Sarah M. and SORANNO, Patricia A., 2019. Spatially Constrained Spectral Clustering Algorithms for Region Delineation. *arXiv.org* [online]. 21 May 2019. [Accessed 19 February 2020].

Available from: https://arxiv.org/abs/1905.08451v1

ZHANG, Jingyu, CHEN, Hengyu, LI, Ruoyan, TAFT, David A., YAO, Guang, BAI, Fan and XING, Jianhua, 2019. Spatial clustering and common regulatory elements correlate with coordinated gene expression. *PLOS Computational Biology*. 1 March 2019. Vol. 15, no. 3, p. e1006786. DOI 10.1371/journal.pcbi.1006786.

ZHENG, Wei, ZHU, Xiaofeng, WEN, Guoqiu, ZHU, Yonghua, YU, Hao and GAN, Jiangzhang, 2020. Unsupervised feature selection by self-paced learning regularization. *Pattern Recognition Letters*. 1 April 2020. Vol. 132, p. 4–11. DOI 10.1016/j.patrec.2018.06.029.

ZHU, Pengfei, XU, Qian, HU, Qinghua and ZHANG, Changqing, 2018. Co-regularized unsupervised feature selection. *Neurocomputing*. 31 January 2018. Vol. 275, p. 2855–2863. DOI 10.1016/j.neucom.2017.11.061.

ZHU, Pengfei, ZHU, Wencheng, HU, Qinghua, ZHANG, Changqing and ZUO, Wangmeng, 2017. Subspace clustering guided unsupervised feature selection. *Pattern Recognition*. 1 June 2017. Vol. 66, p. 364–374. DOI 10.1016/j.patcog.2017.01.016.

ANEXOS

Historias de Usuario

Historia de Usuario: “Importar rasgos temáticos”	
Número: 1	Nombre HU: Importar rasgos temáticos
Usuario: Experto	
Prioridad en Negocio: Alto	Riesgo en Desarrollo: Alto
Puntos Estimados: 1	Iteración Asignada: 1
Programador Responsable: Monica Frómeta Torres	
Descripción: El método debe ser capaz de importar los rasgos temáticos correspondientes a cada capa de la base cartográfica desde diferentes fuentes.	
Observaciones:	

Historia de Usuario: “Obtener rasgos geoespaciales a través de QGis”	
Número: 2	Nombre HU: Obtener rasgos geoespaciales a través de QGis
Usuario: Experto	
Prioridad en Negocio: Alto	Riesgo en Desarrollo: Alto
Puntos Estimados: 1	Iteración Asignada: 1
Programador Responsable: Monica Frómeta Torres	
Descripción: El método debe ser capaz de obtener los rasgos geoespaciales a través de QGis.	
Observaciones:	

Historia de Usuario: “Construir grafo de restricciones espaciales”	
Número: 3	Nombre HU: Construir grafo de restricciones espaciales
Usuario: Experto	
Prioridad en Negocio: Alto	Riesgo en Desarrollo: Alto
Puntos Estimados: 1	Iteración Asignada: 2
Programador Responsable: Monica Frómeta Torres	
Descripción: El método debe construir el grafo de restricciones espaciales a partir de: • los rasgos geoespaciales • los rasgos temáticos.	
Observaciones:	

Historia de Usuario: “Construir ICLA”	
Número: 4	Nombre HU: Construir ICLA
Usuario: Experto	
Prioridad en Negocio: Alto	Riesgo en Desarrollo: Alto
Puntos Estimados: ½	Iteración Asignada: 3
Programador Responsable: Monica Frómeta Torres	
Descripción: El método debe construir el ICLA a partir de: • grafo de restricciones espaciales	
Observaciones:	

Historia de Usuario: “Inicializar el ICLA”	
Número: 4.1	Nombre HU: Inicializar el ICLA
Usuario: Experto	
Prioridad en Negocio: Alto	Riesgo en Desarrollo: Alto
Puntos Estimados: ½	Iteración Asignada: 3
Programador Responsable: Monica Frómeta Torres	
Descripción: El método debe inicializar el ICLA creado.	
Observaciones:	

Historia de Usuario: “Generar subconjunto”	
Número: 5	Nombre HU: Generar subconjunto
Usuario: Experto	
Prioridad en Negocio: Alto	Riesgo en Desarrollo: Alto
Puntos Estimados: ½	Iteración Asignada: 4
Programador Responsable: Monica Frómeta Torres	
Descripción: El método debe ser capaz de generar un subconjunto de rasgos mediante la evolución del ICLA.	
Observaciones:	

Historia de Usuario: “Evaluar subconjunto”	
Número: 6	Nombre HU: Evaluar subconjunto
Usuario: Experto	
Prioridad en Negocio: Alto	Riesgo en Desarrollo: Alto
Puntos Estimados: ½	Iteración Asignada: 4
Programador Responsable: Monica Frómeta Torres	
Descripción: El método debe ser capaz de evaluar el subconjunto de rasgos obtenido a través de: • medida de evaluación a optimizar (maximizar)	
Observaciones:	

Historia de Usuario: “Construir regionalización”	
Número: 7	Nombre HU: Construir regionalización
Usuario: Experto	
Prioridad en Negocio: Alto	Riesgo en Desarrollo: Alto
Puntos Estimados: 1	Iteración Asignada: 5
Programador Responsable: Monica Frómeta Torres	
Descripción: El método debe ser capaz de construir la regionalización a través de: • subconjunto de rasgos óptimos obtenidos.	
Observaciones:	

Historia de Usuario: “Eliminar regionalización”	
Número: 8.1	Nombre HU: Eliminar regionalización
Usuario: Experto	
Prioridad en Negocio: Media	Riesgo en Desarrollo: Bajo
Puntos Estimados: ½	Iteración Asignada: 5
Programador Responsable: Monica Frómeta Torres	
Descripción: El método debe ser capaz de eliminar cualquier regionalización creada.	
Observaciones:	

Historia de Usuario: “Visualizar regionalización”	
Número: 8.2	Nombre HU: Visualizar regionalización
Usuario: Experto	
Prioridad en Negocio: Alto	Riesgo en Desarrollo: Bajo
Puntos Estimados: ½	Iteración Asignada: 5
Programador Responsable: Monica Frómeta Torres	
Descripción: El método debe ser capaz de visualizar la regionalización creada.	
Observaciones:	

Historia de Usuario: “Exportar regionalización como imagen”	
Número: 9	Nombre HU: Exportar regionalización como imagen
Usuario: Experto	
Prioridad en Negocio: Alto	Riesgo en Desarrollo: Medio
Puntos Estimados: 1	Iteración Asignada: 6
Programador Responsable: Monica Frómeta Torres	
Descripción: El método debe ser capaz de exportar la regionalización creada como una imagen.	
Observaciones:	

Historia de Usuario: “Exportar regionalización hacia una hoja de cálculo”	
Número: 10	Nombre HU: Exportar regionalización hacia una hoja de cálculo
Usuario: Experto	
Prioridad en Negocio: Alto	Riesgo en Desarrollo: Alto
Puntos Estimados: 1	Iteración Asignada: 6
Programador Responsable: Monica Frómeta Torres	
Descripción: El método debe ser capaz de exportar la regionalización creada hacia una hoja de cálculo.	
Observaciones:	