



Propuesta de mejora de los circuitos educativos utilizando Clustering y Sistemas de Información Geográfica (SIG)

Olea, Rubén¹; Araujo, Pedro¹

¹ Grupo de Investigación en Tecnologías Informáticas Avanzadas,
Facultad Regional Tucumán, Universidad Tecnológica Nacional
Rivadavia 1050, San Miguel de Tucumán
rubenoolea@gmail.com, pedro.araujo@gitia.org

Resumen. La evolución, a través del tiempo, del sistema educativo argentino hacia la universalización implicó llevar a cabo transformaciones en lo organizativo, hubo que generar sistemas de administración de grandes contingentes de alumnos, docentes y recursos necesarios. El sistema educativo está fuertemente ligado a lo territorial, ya que la distribución espacial de la población es heterogénea y se encuentra estrechamente relacionada con factores de distinta índole. La provincia de Tucumán no está exenta de esta heterogeneidad en la distribución territorial de la población, ni sus organismos de gobierno, incluido el sistema educativo, en este contexto el Ministerio de Educación (ME) de la Provincia de Tucumán divide a su jurisdicción territorial en: a) Zonas para calcular la Bonificación por ubicación, b) Zonas de supervisión y c) Circuitos territoriales. Esta última tiene una lógica basada en los límites administrativos. Dentro de las técnicas de aprendizaje no supervisado encontramos el clustering, cuya aplicación nos permite identificar grupos de elementos similares entre sí a partir de un universo de observaciones con diferentes características o propiedades. En el clustering no se conoce con antelación la estructura de las categorías resultantes. En este trabajo, a través de la aplicación de Clustering y SIG, proponemos como mejorar la división en circuitos territoriales.

Palabras Clave: Clustering, Sistema de información geográfica, SIG, Educación, Tucumán, Circuito educativo, Zonificación territorial.

1. Introducción

La República Argentina, a lo largo de su historia y especialmente en el comienzo de este nuevo milenio, avanzó en la definición de normativas que garanticen y mejoren el derecho a la educación de toda su población. La Ley de Educación Común N° 1.420 (año 1884) definió la escuela primaria como obligatoria y gratuita. La evolución de la matrícula primaria creció junto a la población a lo largo del tiempo. En cambio, la matrícula secundaria crece a través de saltos respecto del tamaño de la población, dándose el primero durante el peronismo, el segundo a partir del año 1983 con el retorno de la democracia. Otro salto se da con la sanción de la Ley Federal de Educación, N° 24.195 (año 1993), que establece la obligatoriedad de 10 años de escolaridad. En el año 2006, la sanción de la Ley de Educación Nacional No 26.206, extiende la obligatoriedad de la educación a toda la secundaria, y establece la universalización de la sala de cuatro años del nivel inicial [1]. El decreto No 1602/09 crea la Asignación Universal por Hijo, el cual “ha fomentado el acceso a los servicios de salud y la escolarización de sus beneficiarios, especialmente en el nivel secundario. Varios autores destacan que esta política ha significado un aumento considerable de la matrícula escolar” [2].

Esta evolución hacia la universalización implica profundas transformaciones. “Para la política educativa esto supone una disminución sensible de la profundidad en favor de la extensión” [3]. Sobre todo, estos cambios deben ser llevados a cabo desde lo organizativo a nivel macro, ya que habrá que generar sistemas de



administración de inmensos contingentes de alumnos, docentes y escuelas. Esto implica tanto el crecimiento de la inversión (logrado en el año 2005 con la Ley de Financiamiento Educativo N° 26.075), como la capacidad organizacional y del planeamiento del sistema educativo.

A pesar de los resultados obtenidos para lograr la universalización de la educación, el sistema educativo se encuentra en un proceso de crisis que afecta a muchos de los aspectos que durante décadas no fueron cuestionados [4]. La educación debe adaptarse a la situación de crisis, donde los beneficiarios de dicho derecho necesitan conocimiento y herramientas. “Las crisis económicas involucran dinámicas de exclusión, concentración y presión sobre los recursos” [5]. A su vez, la educación está fuertemente ligada a lo territorial, ya que su tarea en sus relaciones con el territorio, es la de movilizar todo su talento humano, social e intelectual para que el sujeto-ciudadano entre al mundo y se sostenga en el [6]. La permanencia de los alumnos en el sistema educativo está fuertemente vinculada a diversos factores, entre los que podemos identificar las condiciones socio-económicas del territorio en el que habitan, las características del mercado laboral en el que están insertos los adultos, el acceso a los servicios básicos como salud e infraestructura, acceso a las tecnologías de la comunicación, entre otras.

La distribución espacial de la población es heterogénea, y está estrechamente relacionada con factores de carácter geográfico, económico, social, político, ambiental, cultural e histórico [7]. Por ello, podría decirse que la concentración de población en determinadas regiones no es más que una reacción a una mayor disponibilidad de recursos, infraestructura y servicios, así como también la dispersión de población responde a la falta de los factores indicados.

El gobierno del sistema educativo se ejerce a través de los organismos centrales de decisión, y también mediante el accionar de niveles intermedios que se ubican entre las instancias de toma de decisiones y las escuelas, y que involucran múltiples actores. Es así que, la descentralización se constituye en la estrategia política que permitirá la representación y presencia de un Estado cercano a las realidades locales. La descentralización, entonces, trae consigo connotaciones en términos de toma de decisiones oportunas, adecuadas y contextualizadas.

El artículo está organizado de la siguiente manera: en la Sección 2 se presenta el contexto actual de los circuitos educativos. La Sección 3 se realiza una detallada descripción del problema actual y sus necesidades. La Sección 4 se da introducción a los sistemas de información geográfico (SIG), que serán la base del presente trabajo. En la Sección 5 se describe la metodología utilizada para la resolución del problema presentado. En la Sección 6 se detalla cuáles fueron los resultados obtenidos y su respectivo análisis. Por último, en la Sección 7 se presentan las conclusiones del artículo y los trabajos futuros.

2. Contexto Actual

La provincia de Tucumán no está exenta de esta heterogeneidad en la distribución territorial de la población, ni los organismos de gobierno lo están de tener que organizarse territorialmente en función de sus obligaciones y necesidades. Es por ello que el ME para su normal funcionamiento administrativo-pedagógico divide a su jurisdicción según tres distintos criterios, a saber:

- En zonas definidas a partir de la Resolución N° 217/1991 del Consejo de Educación, que define los parámetros para calcular la Bonificación por Ubicación (adicional salarial de los docentes conocido como zona desfavorable).
- En zonas de supervisión, definida por la necesidad de supervisar pedagógicamente el funcionamiento de los establecimientos educativos.
- En circuitos territoriales, los cuales están definidos con la lógica de permitir a la población docente, e indirectamente a la población en general, de cada circuito desarrollarse íntegramente dentro del mismo.

Estos distintos niveles de división territorial implementados, no fueron definidos en un mismo momento, ni

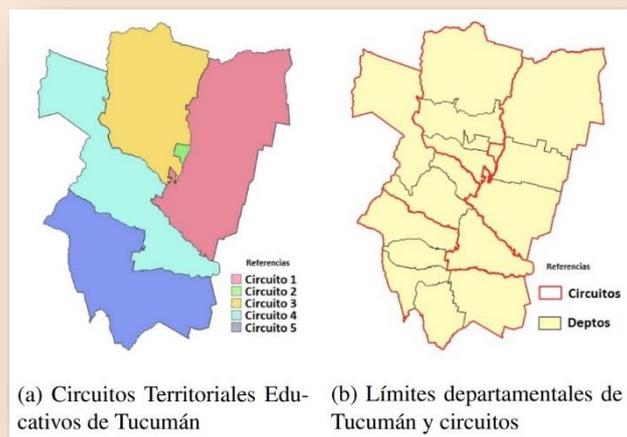
con criterios similares. La división territorial en función de la bonificación por ubicación, nace a partir de una normativa interna del ME, intenta equiparar desde lo económico las distintas realidades en las cuales está espacialmente inserta cada escuela. La división territorial en zonas de supervisión de nivel primario se estableció por una necesidad natural del sistema, referida al control interno del mismo. No se encontró marco normativo que identifique las características territoriales de esta división, por lo que podemos decir que se efectivizó por uso y costumbre. Hay una división territorial de supervisión para cada nivel educativo. La tercer división, la territorial, fue la última en implementarse. Y su aplicación obligó, en forma muy escasa, el replanteo de algunas zonas de supervisión, para evitar el solapamiento. En otras palabras, que una misma supervisión no se intersecte espacialmente con más de un circuito. Los circuitos territoriales fueron definidos tomando como base la división política departamental de la Provincia de Tucumán. Es por ello que no se amolda a la zonificación según supervisión, obligando a realizar pequeños ajustes a lo ya existente.

El sistema educativo tucumano cuenta con: a) 485.899 alumnos (tomando alumnos de Educación Formal y de Educación no Formal) [8] b) 1.990 establecimientos educativos (incluyendo centros unipersonales) [9] c) 36.548 personas conforman el plantel de docentes, directivos, administrativos, auxiliares y otros [10]; d) 1.264 localizaciones escolares [9]. Estos datos corresponden a todos los niveles, tipos y modalidades de educación, tanto del sector estatal como privado, y dentro de lo estatal se encuentran los establecimientos dependientes de la provincia, de los municipios, y de las universidades.

El objetivo del presente trabajo es proponer una mejora en la definición de los circuitos territoriales. La misma se llevará a cabo utilizando técnicas de SIG y distintos algoritmos de agrupamientos.

3. Descripción del Problema

A través de la Resolución N° 753/5 (Med), y su modificatoria la N° 1.070, el ME implementa los *circuitos territoriales* (Figura 1a). Esta división territorial propone posibilitar a los aspirantes (docentes) a ocupar cargos, el acceder a mayores oportunidades laborales en relación con su residencia [11] [12]. Los circuitos educativos complementan la división territorial en zonas de supervisión existentes previamente en el sistema educativo.



Desde una perspectiva multidimensional esta organización promueve la articulación de recursos, capacidades y potencialidades, también posibilita esquemas de cogestión entre los distintos niveles de ejecución del sistema educativo y los organismos de gestión institucional o local [13] La aplicación de los circuitos educativos tuvo como fundamento principal el evidenciar una mayor calidad, integralidad, flexibilidad y



relevancia de las ofertas educativas, cuando en su diseño y ejecución participan los actores del sistema educativo y otros como Salud, Desarrollo Social, trabajo, Empleo, Producción, etc. [13].

Como se observa en las Figuras 1a y 1b esta división se corresponde íntegramente con los límites administrativos departamentales de la provincia de Tucumán.

Sin embargo, esta división territorial tiene un potencial de ser mejorada. Tucumán, en cuanto a su territorio, es la provincia más pequeña en territorio y la más densamente poblada [14], o es la segunda más pequeña y segunda más densamente poblada si se considera a la Ciudad Autónoma de Buenos Aires [14], fisiográficamente, podemos decir que Tucumán se encuentra formada por tres unidades de relieve principales: llanura, montañas y cuencas y valles inter e intramontanos [15]. La llanura ocupa el sector oriental de la provincia, la región montañosa ocupa la porción oeste, con dirección predominante norte-sur, integrada por varios sistemas montañosos que ocupan casi la mitad oeste y las Sierras del Sudoeste tucumano. Entre estos se puede distinguir la tercera unidad fisiográfica, integrada por los valles calchaquíes el valle de Tafi y el valle de la zona del Embalse de Escaba (todos en la zona montañosa), y la cuenca Tapia-Trancas (al norte) [16]. La distribución territorial de la población no es uniforme, el Departamento Capital (en el corazón de la provincia) concentra en 90 km², más del 40% de la población total de Tucumán, el 65% de los habitantes vive en el área del Gran San Miguel de Tucumán y Tafi Viejo (GSMT), que incluye el Dpto. Capital y ciudades aledañas, mientras que departamentos como Tafi del Valle, Trancas, Burruyacú (al oeste, norte y noreste de la provincia respectivamente) tienen baja densidad de población (lo que coincide con su bajo grado de urbanización), indicando una gran dispersión de la población [16]. La planta docente del sistema educativo está implícitamente incorporada en esta descripción. Los docentes se inscriben para ocupar cargos a través de la junta de clasificación. Pudiendo postularse, según sus preferencias personales, a las vacancias de los establecimientos educativos. En esta selección inciden diversos factores, entre los que debemos mencionar lo emocional, el bienestar físico, lo económico y la percepción de la realidad territorial en donde se encuentra el establecimiento poseedor de la vacante a ocupar. En las áreas rurales los docentes tienen que recorrer grandes distancias, por lo general en sus vehículos particulares que muchas veces no están preparados para esos terrenos. Esto hace que las escuelas rurales sean un destino laboral que no siempre es aceptado por los docentes y presentan alta rotación de personal [17].

Observando la actual división territorial, no hay oferta de Educación Superior No Universitaria (SNU) en la ciudad de Yerba Buena (según matrícula 2019 [9]), sin embargo, esta ciudad es considerada entre las que más creció en términos socio-espaciales en este último tiempo, ubicándose entre los tres centros más poblados del GSMT [18]. Yerba Buena funcionó, hasta hace poco, como un componente residencial, de modo que los habitantes se veían obligados a trasladarse en forma pendular, en busca de servicios, básicos como educación y salud o también hacia sus lugares de trabajo, todos centralizados en la ciudad de San Miguel de Tucumán (SMT) [19]. Una consecuencia de esto es la no existencia de SNU por parte del Estado, sin embargo, recientemente se instaló una oferta universitaria desde lo privado. El mirar a San Miguel originó, entre otras cosas, que los medios de transporte y las vías de comunicación no permitan una eficiente comunicación con otras localidades como Tafi Viejo, o Lules. El circuito IV incluye la zona de los valles calchaquíes (Valle de Tafi y Valle de Santa María) y la llanura tucumana. La ruta provincial 307 puede considerarse como la única vía de comunicación entre estas dos regiones geográficas. Las vías de comunicación de la zona de llanura con los valles calchaquíes, en territorio tucumano son escasas o casi nulas, a excepción de la Ruta 307. La opción de cierre o corte de esta vía obliga a tener en cuenta dos vías de comunicación alternativas, una por la provincia de Salta y otra por la provincia de Catamarca.

4. Sistemas de Información Geográfica

En un sentido genérico, los SIG son herramientas que permiten a los usuarios crear consultas interactivas, analizar la información espacial, y presentar los resultados de todas estas operaciones. En el sentido más estricto, es cualquier sistema de información capaz de integrar, almacenar, editar, analizar, compartir y mostrar la información geográficamente referenciada. Los SIG como herramienta, permiten resolver



problemáticas con una fuerte componente espacial, lo que nos ayuda a lograr una mayor eficiencia, en el estudio de las problemáticas vinculadas a lo territorial. Con lo anterior, una definición más precisa es decir que un SIG es un sistema que integra tecnología informática, personas e información geográfica, y cuya principal función es capturar, analizar, almacenar, editar y representar datos georreferenciados. Los SIG permiten la integración de números ilimitados de variables, representando cada uno de ellas un tema espacial (red de caminos, hidrografía, ocupación del suelo, etc.). Cada una de estas variables recibe el nombre de capa temática o layer (es frecuente encontrarlo con su denominación en inglés layer), pudiéndose utilizar diferentes algoritmos para llevar a cabo geoprocursos entre ellas. %La Figura 1 muestra el concepto de capa temática o layer en SIG.

Un SIG posee una estructura compuesta por: 1) datos: la materia prima con la que podemos construir productos de información o mapas que nos ayuden a hacer nuestros análisis y tomar las decisiones en nuestra organización. De ellos dependen, por ejemplo, los procesos que podremos o no realizar con los mismos, y en general todo cuanto podemos esperar de ellos; 2), análisis, métodos y procesos enfocados al análisis de los datos: el análisis es una de las razones fundamentales que llevaron al desarrollo de los SIG. Un SIG siempre incorpora una serie de formulaciones que permiten la obtención de resultados y el análisis de los datos espaciales. Las formulaciones representan procesos que pueden ser sencillos o sumamente complejos, y que pueden resultar de aplicación en otros campos. Su origen puede no ser del ámbito puro de la geografía, por ejemplo, la estadística aporta a los SIG muchas de sus ideas constituyendo un nuevo conjunto de procesos de análisis; 3) visualización, métodos y fundamentos relacionados con la representación de los datos: no es solo una forma más de trabajar con los datos, sino que es la principal. La información geográfica tiene una inherente naturaleza visual, ya que el ser humano entiende al espacio en forma gráfica; 4) tecnología SIG: se incluye aquí tanto el hardware como el software, los cuales conforman un binomio donde encontraremos distintas alternativas, y que se enriquece con la evolución de la tecnología.; 5) factor organizativo: engloba los elementos relativos a la coordinación entre personas, datos y tecnología, o la comunicación entre ellos, entre otros aspectos. Como ejemplo, respecto de los usuarios tenemos distintos perfiles, desde data enters, analistas, tomadores de decisión, y con la aparición de la WEB 2.0 y el bajo costo de las unidades GPS, los SIG han llegado a usuarios no especializados. Resulta evidente el gestionar correctamente la complejidad de los SIG [20].

La primera aplicación del concepto fue en 1832 cuando Charles Picquet creó un mapa que representaba un brote de cólera en 48 distritos de París. Este mapa fue una versión temprana de un mapa de calor. Inspirado por Picquet, John Snow adoptó el mismo principio para registrar en un mapa las muertes por cólera en Londres en 1854. %Fig. 2. Identificó así, que las personas que bebían agua proveniente de bomba de Broad Street, tuvieron una alta tasa de mortalidad. Procediendo a la clausura de la misma, logrando disminuir la mortandad en el barrio. Si bien la cartografía ya existía, el mapa de John Snow fue el primero que, no sólo representó la realidad, sino que permitió analizar fenómenos espaciales dependientes.

5. Metodología

La geografía tiene un rol importante en las sociedades humanas, incide plenamente en la vida humana. En el campo de la geografía económica, se condensa el comportamiento humano frente a las instituciones regionales, ya que intenta explicar la relación entre la geografía y las instituciones.

La humanidad a lo largo de la historia segmentó y delimitó el territorio de diversas maneras para apoyar sus actividades administrativas, políticas y económicas. Por lo general, la mayoría de los límites administrativos de distintas organizaciones, aun dentro de un mismo estado, se han definido de una manera descoordinada. Esto es porque las organizaciones particionan el territorio para satisfacer sus propias necesidades. Se generan así capas, que representan esta realidad, que no permiten realizar análisis con precisión, aún con la tecnología más sofisticada de los SIG [21].

En la literatura se encontró que los métodos de agregación espacial han recibido una gran atención debido a su aplicabilidad a diferentes contextos, estos métodos pueden clasificarse como se indica en la Figura 2 [22]. El diseño de zonas es un problema geográfico que está presente en un amplio espectro de aplicaciones, por

ejemplo, la definición de distritos políticos-electorales, la definición de zonas para prestación de servicios, la zonificación para la asignación de usos del suelo, entre otras [23].

Se hace alusión a los métodos de agregación espacial de distinta manera: regionalización, diseño de zonas, zonificación, delimitación de distritos, diseño o demarcación del territorio, asignación de unidades espaciales, “clustering” espacial o geográfico, agregación espacial, clasificación, partición, teselación espacial, etc. [23]. En este artículo nos centramos en los métodos que pueden aplicarse para determinar regiones analíticas. Nos referiremos a ellos como “métodos de regionalización”, para usar un término que engloba la amplia gama de campos en los que estos métodos han sido creados o aplicados. Su variedad hace que sea difícil proporcionar una definición única del problema de regionalización, pero puede enumerarse algunas características que son comunes a cualquier método que se puede utilizar para definir regiones analíticas:

- Todos los métodos unen áreas geográficas en un número predefinido de regiones mientras optimizan un criterio de agregación particular.
- Las áreas dentro de una región deben estar conectadas geográficamente (la restricción de contigüidad espacial).
- El número de regiones debe ser menor o igual al número de áreas.
- Cada área puede asignarse a una y solo una región.
- Cada región debe contener al menos un área

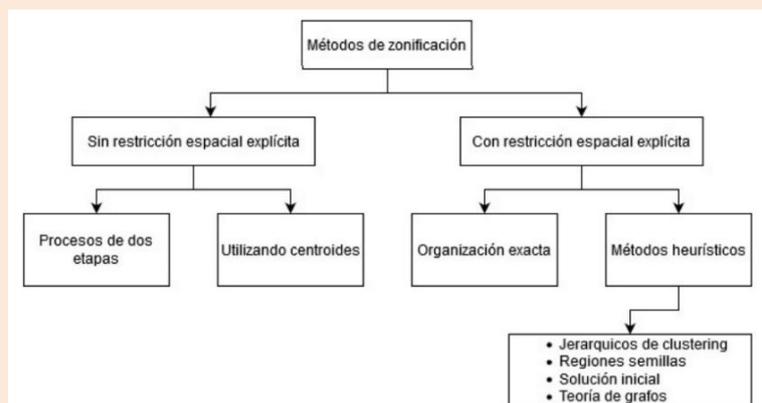


Figura 2: Taxonomía de los métodos para resolver problemas de regionalización [22].

Datos a Utilizar

Los datos a utilizar tienen distintos orígenes y tipologías (puntos, líneas y polígonos). Lo ideal sería utilizar fuentes de datos de igual calidad, como esto no es posible lo común es trabajar con las herramientas que ofrecen los SIG para mejorar los datos en su conjunto. Esta diversidad de fuentes obliga a accionar sobre ellos para procesarlos según determinados criterios con el fin de mejorar su comparabilidad entre sí, y así extraer información. A continuación, se indican los datos básicos que se obtuvieron, la fuente de los mismos.

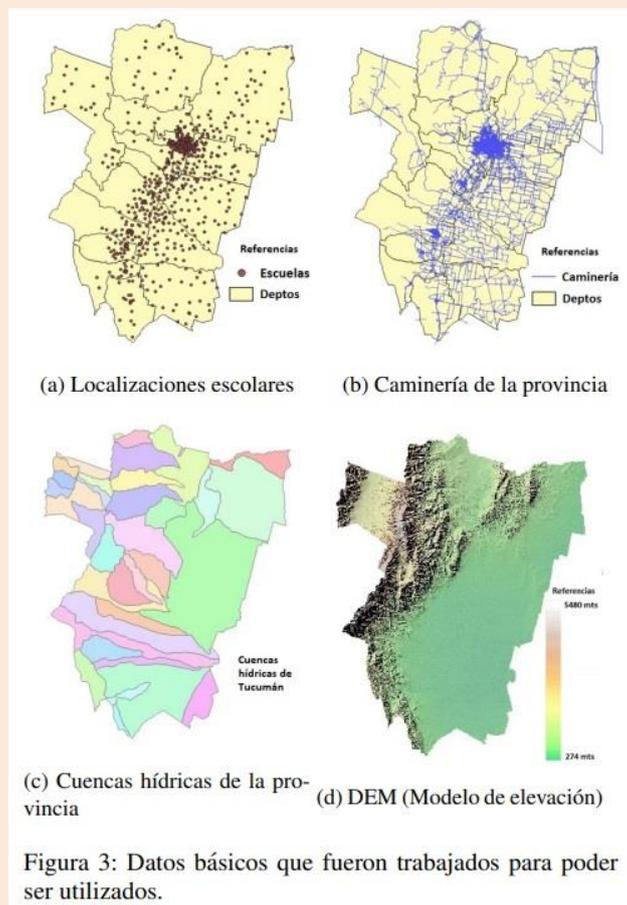
- Ubicación de establecimientos educativos, datos brindados por geoservicios del proyecto Mapa Educativo1 del Ministerio de Educación de la Nación (Figura 3a).
- Listado de establecimientos educativos, publicados por el Dpto. de Estadística Educativa2 del ME.
- Red vial, datos obtenidos del servicio WFS de la Dirección Provincial de Catastro de la Provincia de Tucumán 3 y mejorados con los datos de OpenStreetMap4 (Figura 3b).
- Límites administrativos (departamentales y municipales) de la provincia de Tucumán, datos obtenidos del sitio web del Instituto Geográfico Nacional5 .
- Circuitos educativos, confección propia a partir del mapa publicado por el Dpto. de Estadística

Educativa del ME.

- Cuencas hídricas de la provincia de Tucumán, confección propia a partir del mapa publicado en el sitio web de Red de Información para el Desarrollo Productivo (RIDES)6 (Figura 3c).
- Modelo de elevación, datos de Earth Explorer, del Servicio Geológico de los Estados Unidos (USGS)7 (Figura 3d).

Para cada capa/variable se realizó un trabajo de mejora de la información de distinta índole dado que, en la mayoría de los casos, no se cuenta con los metadatos referidos a la fecha de actualización de cada una. Los metadatos permiten describir el contenido, la calidad, y otras características de los datos. Ayudan a localizar y entender los datos espaciales disponibles

Cada variable se trabajó en función de su tipología y las necesidades de ese dato respecto del problema a resolver. La primera acción sobre los datos recolectados fue llevarlos a un sistema de proyección cartográfico uniforme, se reproyectó la capa utilizando POSGAR 98 Faja 3 (Posgar), según corresponde para la provincia de Tucumán.



Una vez llevados los datos al mismo sistema de proyección, se procedió a trabajarlos según su tipología vectorial/raster y su necesidad dentro del problema. A continuación, se detallan algunos de los trabajos realizados sobre las capas:

Edificios escolares

La capa de establecimientos educativos nos permite visualizar la distribución de la oferta educativa en sus



distintas modalidades. Esta capa fue actualizada con el listado del padrón de establecimientos. Se obtuvo un conjunto de datos que contienen para cada escuela variables como ámbito, localización de la oferta educativa entre otros datos. Este set de datos permite confeccionar distintos temáticos como ubicación de establecimientos rurales o urbanos, establecimientos privados o estatales, establecimientos de dependencia provincial, municipal o nacional, entre otros.

Caminería

Las geometrías que representan la caminería de la provincia, fueron trabajadas desde distintas fuentes. Se completaron los datos para que figuren totalmente las rutas nacionales, y provinciales, se amplió la red terciaria de la caminería. También se trabajó para lograr la integridad del dato, es decir eliminar geometrías repetidas y en garantizar que haya continuidad topológica en las intersecciones de las polilíneas, con el fin de conformar un grafo que permita, a través de la relación de sus componentes, por ejemplo, el calcular las distancias necesarias. Se combinan los datos de las rutas y caminerías terciarias con la red de callejeros de las distintas poblaciones de Tucumán, este trabajo es lento, ya que como se mencionó, se verifica la topología de los datos, y también hay que verificar la correcta definición del sentido de cada tramo de calle o ruta, lo que nos permite determinar el correcto sentido de circulación de las calles. El dato obtenido a partir de la caminería es el que denominaremos distancia, y se corresponde con la distancia existente desde la ubicación del establecimiento educativo hasta la ciudad de SMT o Concepción, siguiendo el trazado de la Caminería. Estas ciudades destino están definidas en la Resolución N° 217 [24].

Circuitos educativos

En el sitio del ME 8 puede observarse como está dividida la provincia de Tucumán en circuitos educativos. También se observa que cada circuito coincide en su totalidad con un departamento o un grupo de departamentos de la provincia. Entonces se replicó, partiendo de la división administrativa en departamentos de la provincia, esta capa.

Elevación

Con la disponibilidad del Modelo de Elevación Digital (DEM) obtenido a partir del proyecto SRTM (Shuttle Radar Topography Mission), descargado del Sitio WEB de USGS, se pudo generar las curvas de nivel de la provincia. De estas se generó una capa con los polígonos correspondientes a cada valor de la escala de las curvas de nivel. Luego se vinculó cada establecimiento o edificio educativo, con los polígonos creados anteriormente, obteniendo de esta forma la cantidad de metros (o elevación) sobre el nivel del mar a los que está ubicado el establecimiento educativo.

Cuencas hídricas

Para obtener este dato se georreferenció la imagen obtenida en el sitio de RIDES. Luego se procedió a replicar en formato vectorial esta imagen, utilizando el software QGIS. Así se obtuvo para cada establecimiento educativo la cuenca hídrica en donde se encuentra ubicada. Como se mencionó, para la mayoría de los datos recopilados no se cuenta con la totalidad de sus respectivos metadatos, los cuales nos ayudan a comprender la calidad de los mismos, ya que permiten describir la información geográfica, transformándose en indispensables para evaluar y analizar los datos. Hay que señalar que la exactitud de los datos espaciales depende directamente del proceso de captura utilizado para obtener dicha información. El determinar el nivel de error de los datos implica, para cada localización o algunas de ellas, tener en cuenta la falta de exactitud de ubicación (como ya se mencionó muy probablemente debido a un tema de escala) y de faltante de datos. El trabajo necesario para la corrección de lo explicitado implica el revisar los datos uno a uno, o al menos hacer un muestreo representativo de los mismos para determinar el nivel de error. Se pueden encontrar distintos tipos de errores en los datos, por ejemplo, cartografía a escalas diferentes, cartografía con diferentes sistemas clasificatorios, problemas debido a la captura en formato digital, capas de igual temática, pero con diferentes sistemas clasificatorios entre otros tipos de errores.

6. Resultados Obtenidos y su Análisis

Durante la ejecución de este proyecto se han utilizado numerosas alternativas de los algoritmos de agrupamientos (clustering) cada uno con sus características. El primero en ser utilizado fue el algoritmo basado en densidad denominado DBSCAN. Este propone que, para incorporar un punto en un agrupamiento, tienen que existir un mínimo de observaciones dentro de un radio de proximidad pequeño, entonces ~ cada agrupamiento obtenido estará separado por espacios vacíos o pocas observaciones. Este método encuentra los agrupamientos que se conforman según los parámetros con que se corre el algoritmo. El inconveniente de utilizar este método es que no se puede indicar la cantidad de agrupamientos deseados, ya que el algoritmo mismo los detecta.

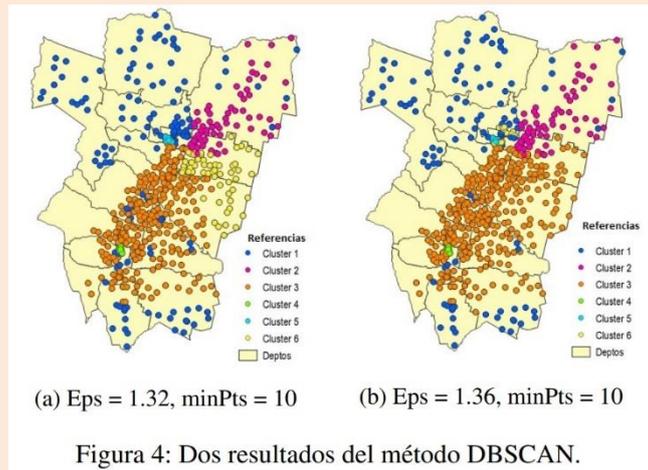


Figura 4: Dos resultados del método DBSCAN.

Las características de la dispersión del universo de observaciones (escuelas) que encontramos en la provincia de Tucumán, hace que este algoritmo tienda a agrupar los puntos de la zona pedemontana en un único clúster. La ciudad de SMT, por la densidad de locaciones que posee, sería parte de este clúster. Por esta razón, previamente eliminamos estos puntos previendo este comportamiento del método.

En las Figuras 4a y 4b se pueden observar los resultados de algunas ejecuciones del algoritmo, siempre obteniendo 5 (cinco) agrupamientos. Los puntos azules corresponden a los puntos ruido, que no son considerados en los agrupamientos del resultado.

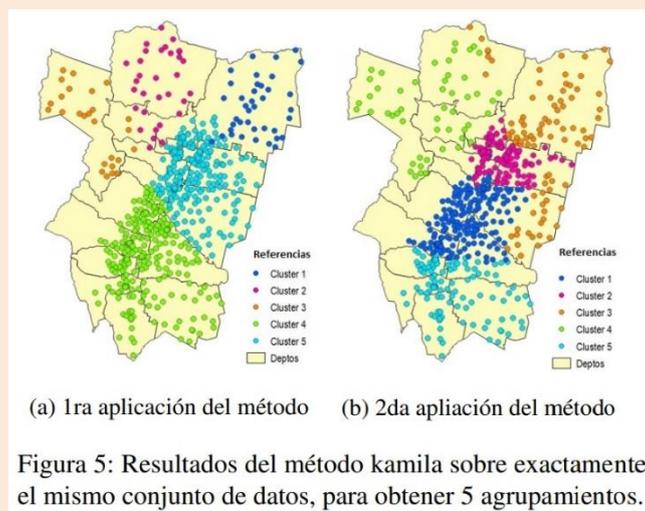


Figura 5: Resultados del método kamila sobre exactamente el mismo conjunto de datos, para obtener 5 agrupamientos.

El segundo algoritmo utilizado fue KAMILA [25]. Este método de agrupamiento equilibra equitativamente la contribución de variables continuas y categóricas. Su utilización ha permitido mejorar los resultados

utilizando la totalidad de los datos con que se cuenta. Se ha trabajado sobre el conjunto de datos con el paquete RStudio, los resultados obtenidos se visualizaron en QGIS9.

Al observar los resultados parecen prometedores, podemos apreciar que los clústeres obtenidos son más uniformes, encontramos una mejor partición del territorio. Sin embargo, cada vez que se ejecuta el método, con los mismos datos de entrada, los resultados que arroja son distintos (Figuras 5a y 5b) constituyendo una de las debilidades más importantes del mencionado algoritmo. Por lo mencionado, la utilización del algoritmo no representa una opción válida, dado que ante la aparición de nuevos establecimientos educativos el mismo daría resultados completamente diferentes a los anteriores.

Por último, y no por ello menos importante, se utilizó el algoritmo K-Means. Este método de agrupamiento partición a un conjunto de elementos en k grupos, en donde cada observación (en este caso los establecimientos) pertenece al grupo cuyo valor medio es más cercano. Se trabajó con este algoritmo para obtener 5, 6, 7, 8, 9 y 10 clústeres. Las variables utilizadas fueron ámbito, departamento, municipio, cuenca hídrica, elevación, distancia, latitud y longitud. Los resultados obtenidos no fueron satisfactorios, ya que, si bien se observa una agrupación espacial, muchos puntos de distintos clústeres se encuentran dispersos por el territorio (Figuras 6a y 6b).

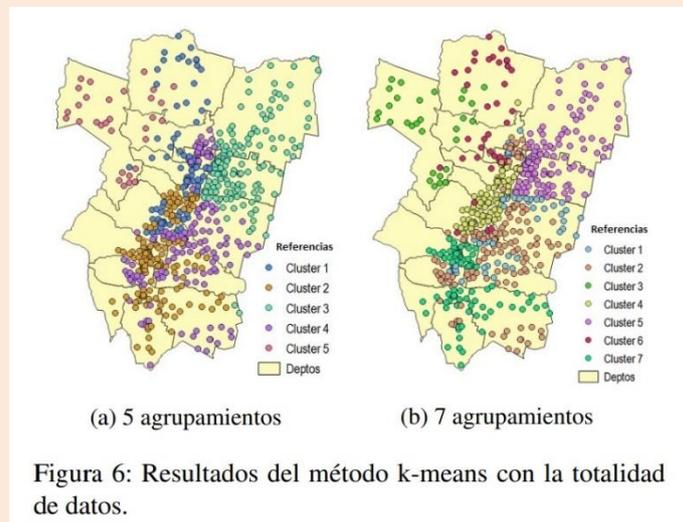


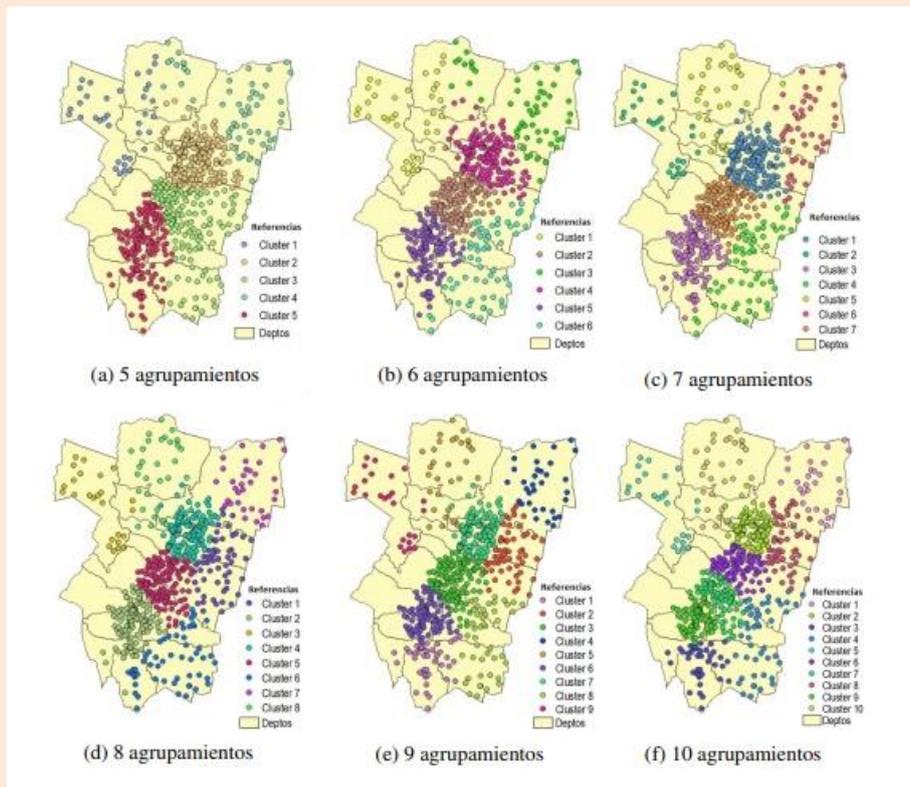
Figura 6: Resultados del método k-means con la totalidad de datos.

Se realizaron a posteriori diversas corridas con el objeto de identificar variables que incidan en esta dispersión de los clústeres en el plano. Primero se descartaron las variables municipio, cuenca y ámbito, ejecutando nuevamente el algoritmo con el resto de las variables. Los clústeres obtenidos son más compactos, no hay tanta dispersión en el resultado. Sin embargo se observa un sesgo que se puede vincular con la variable departamento, por ejemplo el departamento Chichligasta -ubicado en el sudoeste de la provincia de Tucumán- figura vinculado o incluido en su totalidad en clústeres que no son contiguos o figura como un clúster independiente.

Por último, se ejecutó el algoritmo solamente para las variables elevación, distancia, latitud y longitud, dejando de lado las variables ámbito, departamento, municipio, cuenca y zona. El resultado obtenido puede indicarse como aceptable dado que los agrupamientos son bastantes consistentes y bien definidos entre sí (Figuras 7a, 7b, 7c, 7d, 7e, 7f.).

El resultado obtenido a partir de las variables elevación, distancia, latitud y longitud, junto al método K-Means son, a nuestro criterio, mas acertivo. A diferencia de los otros métodos utilizados, en K-Means podemos indicar la cantidad de agrupamientos deseados en el resultado (a diferencia de DBSCAN), y el resultado para una ejecución del método con los mismos datos de entrada, es siempre el mismo (a diferencia de KAMILA).

Observando los resultados del algoritmo K-Means con las variables elevación, distancia, latitud y longitud, advertimos un patrón en los resultados. Se generan cuatro agrupamientos que se mantienen independientemente de la cantidad de clústeres que se obtienen. El primero se encuentra en el oeste de la provincia, el segundo en el corazón de la provincia incluyendo la zona del GSMT, el tercero al sur del anterior, el mismo incluye la ciudad de Concepción, el cuarto agrupamiento se encuentra entre estos dos últimos mencionados.



A medida que se incrementa la cantidad de clústeres a obtener, se mantienen estos agrupamientos principales, pero decreciendo en tamaño a medida que se agregan nuevos ~ agrupamientos. Se percibe una relación entre este decrecimiento y las variables caminaria y elevación, puesto que la variable distancia comienza a dominar en la aparición de nuevos agrupamientos, y la variable elevación mantiene un escalonamiento según sus valores.

7. Conclusiones y Trabajo Futuro

En el presente trabajo se ha propuesto una mejora en la distribución de establecimientos educativos de los circuitos educativos de la Provincia de Tucumán. Si bien, existen en la literatura numerosos trabajos sobre regionalización, no encontramos trabajos basados en la división territorial dedicados a la educación, con criterios emanados por el Ministerio de Educación de la Provincia.

Para alcanzar nuestro objetivo se utilizaron tres algoritmos de agrupamiento largamente conocidos en la comunidad científica, y estos son: DBSCAN, K-Means y KAMILA. Los dos primeros son los algoritmos más conocidos y utilizados en la literatura científica, el tercero fue utilizado porque está adaptado para trabajar con datos de tipo mixto (categóricos y continuos). Cada uno de dichos algoritmos fueron ejecutados con diferentes criterios para lograr una óptima distribución de los circuitos educativos. De los mencionados, solo el último



arrojó resultados, que además de parecernos los más correctos, son posible de ser replicados sin muchos inconvenientes. En este sentido, se realizaron pruebas que van desde buscar la solución con cinco clústeres hasta un máximo 10, dado que un número mayor podría dificultar la gestión administrativa de los mismos por parte del Ministerio de Educación Provincial.

Así, consideramos que la solución está dada por la conformación de 8 (ocho) clústeres realizando muy pequeños ajustes en algunos de los mismos. Si consideramos la distribución de la oferta educativa, los 8 agrupamientos obtenidos contienen cada uno todo tipo de oferta (educación inicial, primaria, secundaria y superior no universitaria), y no se observa que esta división se vea afectada por la existencia de barreras naturales (por ejemplo, el Río Salí, que en su recorrido en la zona sureste de la provincia es una barrera natural). A su vez, esta división territorial puede organizarse agrupando estos clústeres entre sí para conformar agrupamientos mayores, sin romper la lógica inicial.

Como trabajo futuro, se podrían agregar más variables distintas a las descartadas, por ejemplo, aquellas vinculadas a la accesibilidad territorial (tiempo necesario para recorrer las distancias desde los establecimientos a los puntos que se definan, distancia desde el establecimiento a caminos principales, accesibilidad dependiente de eventos climáticos, etc.). Con estas nuevas variables esperamos poder incluir en la definición de los agrupamientos los accidentes geográficos que pueden servir como límite natural de los mismos (por ejemplo, un río sin puentes que permitan atravesarlo). El contar con estas variables implica realizar un trabajo más oneroso en tiempo, y que excede los plazos para concluir este trabajo. A su vez, se espera también utilizar otras técnicas propias de la inteligencia artificial como las redes neuronales artificiales, sistemas multiagentes, entre otros.

Referencias

- [1] María Catalina Nosiglia. El proceso de sanción y el contenido de la ley de educación nacional n° 26206: continuidades y rupturas. *Praxis educativa*, 11(11):113–138, 2012.
- [2] Santiago Garganta. El Impacto laboral y demográfico de la Asignación Universal por hijo en Argentina. Phd thesis, Facultad de Ciencias Económicas, 2015.
- [3] Axel Rivas, Alejandro Vera, and Pablo Bezem. Radiografía de la educación argentina. Cipeec, 2010.
- [4] Roberto Mur Montero. Título: Notas sobre la organización desconcentrada de las administraciones educativas. *Revista Iberoamericana de Educación*, (3):41–62, 1993.
- [5] Ana Inés Heras Monner Sans and Amalia Miano. Educación, autoorganización y territorio. *Revista mexicana de investigación educativa*, 22(73):533–564, 2017.
- [6] Silvio Cardona González. Educación, territorio y modernización en el contexto de la descentralización del sistema educativo en Colombia en orden al territorio. *Revista Ideales*, 1(1), 2016.
- [7] Rodolfo Montaña Salazar and Gerardo Núñez Medina. Metodología para medir la concentración y dispersión demográfica de la población mexicana a través del modelo policéntrico.
- [8] Ministerio de Educación de Tucumán Dirección de Calidad Educativa. Evolución de la matrícula, Total Provincia - Tucumán, 1999-2019
- [9] Dirección de Calidad Educativa Departamento de Estadística Educativa. Datos de establecimientos - Dirección de Calidad Educativa.
- [10] Dirección Nacional de Información y Estadística DiNIECE. CENPE 2014 Censo Nacional del Personal de los Establecimientos Educativos: Datos generales.
- [11] Ministerio de Educación. Resolución no 753/5 med, September 2009.
- [12] Ministerio de Educación. Resolución no 1070/5 med, October 2011.
- [13] Superior Gobierno de la Provincia de Tucumán. La escuela, un espacio para reponer las palabras. Department of Geoinformation, Technical Univ., 12 2015.
- [14] Instituto Nacional de Estadística y Censos (INDEC). Censo Nacional de Población, Hogares y Viviendas 2010: resultados definitivos. Number 2 in B. INDEC, 1 edition, 2012.
- [15] María E Puchulu and Diego S Fernández. Características y distribución espacial de los suelos de la



provincia de Tucumán. Geología de Tucumán. (Eds S Moyano, ME Puchulu, D Fernández, G Aceñolasa, ME Vides, S Nieva) pp, pages 1–17, 2016.

[16] Patricia Mónica Fernández and Daniel Esteban Quiroga. Aportes metodológico-conceptuales de la distribución espacial de la población: El caso de Tucumán y Catamarca en el noroeste argentino [1991, 2001 y 2010]. In VII Jornadas de Sociología de la UNLP 5 al 7 de diciembre de 2012 La Plata, Argentina. Argentina en el escenario latinoamericano actual: Debates desde las ciencias sociales. Universidad Nacional de La Plata. Facultad de Humanidades y Ciencias de la ..., 2012.

[17] M. Rosa Brumat. Maestros rurales: condiciones de trabajo, formación docente y práctica cotidiana. Revista iberoamericana de educación, 55(4):1–10, 2011. Publisher: Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura.

[18] Matilde Malizia and Paula Boldrini. Las lógicas de ocupación del espacio urbano. un estudio de realidades contrapuestas. el caso de las urbanizaciones cerradas y villas miseria en yerba buena, gran san miguel de Tucumán. Cuadernos de la Facultad de Humanidades y Ciencias Sociales-Universidad Nacional de Jujuy, (41):197–219, 2012.

[19] Alicia Inés Ferrari, Agustín José Machuca, and Sandra L Mansilla. Transformación y articulación del territorio urbano en el gran san miguel de Tucumán: El caso de yerba buena. Breves Contribuciones del Instituto de Estudios Geográficos, (27):89–105, 2016.

[20] Víctor Olaya. Sistemas de información geográfica. CreateSpace, 2016.

[21] Serryn Eagleson, Francisco Escobar, and Ian Williamson. Automating the administration boundary design process using hierarchical spatial reasoning theory and geographical information systems. International Journal of Geographical Information Science, 17(2):99–118, 2003.

[22] Hada Melissa Sáenz Vela. Revisando los métodos de agregación de unidades espaciales: Maup, algoritmos y un breve ejemplo. Estudios demográficos y urbanos, 31(2):385–411, 2016.

[23] Pilar Moreno, Jesús García, and LDE Lacalle. Estado del arte en procesos de zonificación. GeoFocus, (11):155–181, 2011.

[24] Consejo de Educación. Resolución N° 217, March 1991.

[25] Alexander H. Foss and Marianthi Markatou. kamila: Clustering mixed-type data in R and Hadoop. Journal of Statistical Software, 83(13):1–45, 2018.