



Inteligencia Artificial para la Inteligencia Urbana

Sistema de Ayuda a la Decisión para la predicción de la vulnerabilidad social

Francisco Javier Abarca-Alvarez

Departamento de Urbanística y Ordenación del Territorio
Universidad de Granada
Granada, España
fcoabarca@ugr.es

Francisco Sergio Campos-Sánchez

Departamento de Urbanística y Ordenación del Territorio
Universidad de Granada
Granada, España
scampos@ugr.es

Resumen— La vulnerabilidad social desde un punto de vista socio ambiental se enfoca en la identificación de grupos desfavorecidos o vulnerables y de las condiciones de los entornos en los que habita, tratando de determinar los factores que explican la dificultad para afrontar situaciones de desventaja social. Por su complejidad, ligada a la multidimensionalidad, no siempre es sencillo identificar los estratos sociales y urbanos afectados. La investigación pretende evaluar la conexión entre ciertas dimensiones de la vulnerabilidad social y su contexto residencial, creando un Sistema de Ayuda a la Decisión (DSS), útil para la planificación de acciones sociales y urbanas. Para ello se realiza una aproximación holística a la información censal y demográfica habitual en estos estudios, proponiendo la construcción de (i) un modelo de conocimiento basado en Redes Neuronales Artificiales (ANN) del tipo Mapa Auto-organizado (SOM) con el que se identifica y caracteriza un perfil demográfico susceptible de presentar vulnerabilidad social, y (ii) un modelo predictivo de tal perfil basado en reglas a partir de variables residenciales, construido mediante árboles de inferencia condicional. Tales modelos combinados con Sistemas de Información Geográfica (SIG) hacen viable un DSS para la predicción de la Vulnerabilidad Social a partir de información residencial.

Palabras clave— Vulnerabilidad Social; Mapa Auto-organizado; árbol de decisión condicional; DSS; Smart City.

INTRODUCCIÓN

Se suele definir la vulnerabilidad como el potencial de sufrir pérdidas o daños [1], físicos o económicos y localizados en un territorio específico. Cuando se aborda la vulnerabilidad desde un punto de vista social, se enfoca nítidamente en su vertiente humana de aplicación poblacional [2]. En torno al concepto de vulnerabilidad social han aparecido en los últimos años multitud de líneas de trabajo, unas más vinculadas a los riesgos naturales y desastres [3], [4], otras a los factores ambientales [5] y otras más próximas concepto de pobreza [4]. De un enfoque que consideraba fundamentalmente los acontecimientos naturales, se pasó paulatinamente a uno que admitía que los efectos sobre la población estaban condicionados por su capacidad de mitigación [6], en la capacidad de una persona o comunidad para anticipar, resistir y sobreponerse al impacto de acontecimientos imprevistos [7], abriéndose en este sentido a un enfoque que pone en el centro a las personas, siendo aquellas las que tienen o no capacidad de sobreponerse [8] o adaptarse a las vicisitudes, no

exclusivamente vinculadas a los riesgos ambientales, e incluso incorporando una vertiente espacial [9].

La vulnerabilidad social presenta diversos retos como son la multidimensionalidad [4], [5], [10]–[12],[11], [12] o el hecho de que muchas de las variables o dimensiones a evaluar no son directamente observables [10]. Entre las investigaciones que han tratado de identificar indicadores de la vulnerabilidad social destaca el trabajo de Cutter, et al. en el que incorpora como variables del llamado *SoVI*, toda una serie de indicadores como son factores socio-económicos, de edad, indicadores de desarrollo comercial o industrial, desempleo, indicadores de ruralidad, propiedad residencial, nivel de infraestructuras, nivel de rentas, ocupación, acceso a servicios médicos, factores de género, raza y etnia, estructura familiar, nivel educativo, crecimiento vegetativo, dependencia a servicios sociales y presencia de población con necesidades especiales [5].

Con el reto de realizar un Sistema de Ayuda a la Decisión (DSS) en relación a la vulnerabilidad social, se plantea como herramienta necesaria el uso de un Sistema de Información Geográfica (SIG). Tales sistemas han sido usados asiduamente por gobiernos, investigadores y empresas como herramienta de decisión en las que alcanza cierta repercusión e influencia la dimensión espacial [13]. Los SIG surgieron a finales de la década de 1960 desarrollándose especialmente a partir de los 80 [14], y generalizándose su uso a partir de 1990 y especialmente ante el compromiso de convertir tanto a ciudadanos como urbanistas, en actores importantes de la planificación. Es preciso tener presente que los GIS no siempre han estado preparados para cumplir como un DSS en cuanto que deberían integrar realidades y problemas complejos para el apoyo de determinadas decisiones [15], consolidándose como unos sistemas flexibles y resilientes.

Desde su origen la realidad informacional sobre la que se apoyan estos sistemas ha cambiado rápidamente, proliferando la presencia de datos e información espacial de acceso libre, configurándose el campo de los SIG prácticamente como una disciplina, que algunos llaman GISciencia [16], apoyándose en la idea de una nueva Geografía Cuantitativa, fundamentalmente espacial, tendente a la planificación y gestión. Con el tiempo los SIG han pasado del énfasis en la “S” por los problemas computacionales (60-70s), pasando a dedicarse a la “I” por el interés en la información (80-90s), para a partir de 2000

enfocarse en la “G”, ante una necesidad de interpretación geográfica, materializándose en la “sociedad de la información geográfica” y abriendo una nueva etapa para la historia de la Geografía [17]. Es en este punto es cuando singularmente los GIS se aproximan propiamente al concepto DSS.

Frecuentemente esta nueva Geografía Cuantitativa está elevando la granularidad de los datos geográficos hasta el extremo, en una auténtica “n-dimensionalidad” de los datos [18]. En la mayoría de los casos, según Pragya Agarwal y André Skupin, se han enfocado en el análisis estadístico tradicional para resolver problemas de autocorrelación espacial, dejando otros múltiples ámbitos totalmente por explorar. Algunos de esos espacios están siendo abordados por enfoques emergentes como son la Inteligencia Artificial o las Redes Neuronales Artificiales, el aprendizaje automático (*Machine Learning*) o de forma específica por la Geo-computación. Estas nuevas técnicas y enfoques están propiciando un cambio de paradigma en los DSS, considerándose que en la actualidad pueden ser útiles para la comprensión de la realidad, detección de sus problemas, y en definitiva la formulación de nuevas hipótesis y no solo como instrumento para verificar aquellas previamente establecidas.

Como objetivo principal la investigación se propone evaluar la viabilidad de la construcción de un modelo de vulnerabilidad social, a priori complejo, a partir de un modelo residencial, del que es más sencillo tener información. Para ello se utilizarán técnicas de interpretación de la realidad mediante Inteligencia Artificial y *Machine Learning* apoyados en SIG. De forma específica para la validación de la metodología propuesta se caracterizará la vulnerabilidad social de la población de Andalucía a partir de información residencial en la que reside la misma.

Para alcanzar tal objetivo la investigación se enmarca en los paradigmas de los DSS orientados al conocimiento y orientados a los modelos [19]. El primer paradigma se enfoca en la construcción de un sistema de descubrimiento de conocimiento basado en bases de datos institucionales sobre las cualidades demográficas y sociales de Andalucía (Fase de modelado 1 de la metodología); y en el segundo paradigma (Fase de modelado 2 de la metodología) se enfatiza el acceso, la manipulación y creación de un modelo cuantitativo de la realidad social, orientado a proporcionar apoyo a la decisión, elaborado a partir de la realidad residencial de los territorios en estudio. Los DSS usan datos y parámetros proporcionados por los agentes de decisión para ayudarles a analizar una situación, aunque no tienen por qué ser datos masivos [19]. En nuestro caso el modelo se construye con datos masivos (tanto demográficos como residenciales), pero puede ser usado para la toma de decisiones con una información muy limitada, incluso escasa.

La investigación, tal y como se ha avanzado, consiste en la construcción de dos modelos vinculado con herramientas procedentes de las Tecnologías de la Información [20]:

Descubrimiento de conocimiento y clasificación mediante aprendizaje no supervisado. Los Mapas Auto-organizados (SOM)

En la investigación se recurrirá a técnicas de lo que se viene a conocer como Mapas Auto-organizados, en adelante SOM. Fueron propuestos inicialmente por Teuvo Kohonen [21]–[23]. La metodología SOM es una técnica descubrimiento de conocimiento o de minería de datos consistente en una Red Neuronal Artificial. Procede del campo de conocimiento de la Inteligencia Artificial, habiéndose mostrado muy eficaz y robusta en numerosas disciplinas, presentando diversas capacidades entre las que podemos destacar inicialmente dos: (i) es capaz de mostrar y visualizar la información de partida de forma clara y ordenada, (ii) permite clasificar y por lo tanto etiquetar los sujetos en estudio en clases que no requieren su definición, caracterización o etiquetado nominativo previo (aprendizaje no supervisado). Frente a otras metodologías de descubrimiento de patrones, como por ejemplo el análisis clúster, la metodología SOM, tiene la ventaja de (i) permitir visualizar un gran conjunto de datos estadísticos [24], (ii) mostrar las relaciones topológicas de similitud o de diferencia entre los sujetos en estudio, (iii) ser interpretables gráficamente y (iv) constituir por sí mismo un sistema de conocimiento de ayuda a la decisión para el análisis y visualización de indicadores estadísticos [24].

Mediante estas técnicas se obtendrá por un lado el etiquetado a modo de clases o perfiles de los diferentes fragmentos de territorio andaluz estudiado, atendiendo al análisis multi-variable de los atributos demográficos y sociales estudiados. Se alcanza asimismo un sistema de análisis e interpretación de las clases obtenidas, facilitado por la metodología SOM, materializado en cartografías temáticas de los diferentes atributos incluidos en la red neuronal y en diferentes tablas y datos estadísticos que permiten conocer las características diferenciadoras de cada perfil. Para facilitar su utilización como parte de un DSS se representan tales clases, y en especial la clase o categoría de vulnerabilidad social mediante GIS.

Construcción de modelos predictivos mediante aprendizaje supervisado. Árboles de decisión

Mediante un proceso de *Machine Learning* se obtendrán una serie de reglas que permitan predecir las clases o perfiles que se determinaron con el modelo SOM, utilizando para ello únicamente atributos sobre la realidad residencial de los territorios en estudio. Estas variables residenciales no se tuvieron en consideración en la creación y constitución de la red neuronal SOM ni consecuentemente pudieron afectar o correlacionarse en la definición de los perfiles obtenidos por aquella. Se realiza una aproximación al problema de aprendizaje mediante el paradigma “divide y vencerás”, que al realizarse sobre un conjunto de instancias independientes, conduce naturalmente a un estilo de representación llamado árbol de decisión [25]. En cada nodo del árbol interviene un atributo en particular, comparándose normalmente cada instancia del atributo con el valor de una constante, y generándose normalmente dos ramas atendiendo a las instancias que cumplen o no tal regla. El árbol supone una representación asequible para interpretar y usar en la



predicción de la realidad demográfica y social de un territorio y consecuentemente útil para la toma de decisiones sobre el mismo, usando para ello una información limitada y generalmente de fácil y económica obtención sobre la realidad residencial del lugar en estudio. Asimismo al evaluar el “valor” de los perfiles alcanzados en la Fase 1, en su caracterización espacial mediante SIG, se verifica la utilidad de la metodología propuesta.

MATERIALES Y MÉTODOS

Metodología de investigación

Para la óptima comprensión y para la obtención de los mejores resultados del DSS se siguen las siguientes fases [26]: (i) información y funciones de procesado, (ii) conjuntos de datos, (iii) modelos, y (iv) representaciones visuales.

1) Información y funciones de procesado:

La información utilizada en la investigación procede del Censo de Población 2001 de Andalucía obtenidos por el Instituto de Estadística y Cartografía de Andalucía (IECA). Sobre esta información se ha realizado una intensa preparación de datos con la integración y limpieza de datos, transformación de atributos mediante la creación de indicadores agregados que sintetice forma objetiva y compacta las principales cualidades demográficas de los datos originales. Debido a la robustez de los SOM, no es necesario realizar su tipificación o normalización [27] previamente a la agregación e incorporación al modelo.

a) Instancias: La unidad de territorio sobre la que se obtienen los datos, es la Sección Censal, alcanzándose la totalidad de las 5381 secciones censales de Andalucía, quedando representada la totalidad de la superficie y de población censada en la región andaluza, no realizándose muestreo alguno.

b) Atributos: De los indicadores de Vulnerabilidad Social “SoVI” [5], se incorporan factores socio económicos, de edad, indicadores de desarrollo comercial o industrial (empleo por sectores), desempleo, indicadores de ruralidad (empleo sector agrario), propiedad residencial, nivel de infraestructuras, nivel de rentas (aunque conectado a nivel de ocupación), ocupación, acceso a servicios médicos, y no se incorporan factores de género, raza y etnia, estructura familiar, nivel educativo, crecimiento vegetativo, dependencia a servicios sociales o presencia de población con necesidades especiales. Los atributos que se han usado en la Fase de modelado 2 están compuestos de variables de la dimensión residencial.

2) Conjuntos de datos:

Se opera inicialmente con dos bases de datos desconectadas: una propia de la Fase de modelado 1, con dimensión principalmente demográfica y social y otra propia para la Fase de modelado 2, basada en la dimensión residencial. El funcionamiento será fundamentalmente independiente, conectándose únicamente tras el Modelado 1 para evaluar cómo se ajustan los perfiles de vulnerabilidad social y para la construcción del árbol de decisión.

3) Modelos:

Tal y como se ha avanzado distinguimos entre dos fases de modelado:

a) Fase de modelado 1: Modelo de clasificación y conocimiento. Entre sus objetivos tenemos por un lado, la clasificación y etiquetado de los datos de la dimensión demográfica y social, tal y como se han identificado anteriormente. En esta fase se usará una red neuronal artificial y específicamente un SOM. Esta metodología en cuanto que no supervisada, permite clasificar (propriadamente segmentar) sin atribuir a priori una etiqueta con definiciones y significados previamente atribuidos, siendo útil para reducir la enorme complejidad de los datos [28].

Para facilitar la comprensión de los perfiles obtenidos se caracteriza a cada clúster con sus estadísticos básicos como son la Media, la Desviación Estándar, el Máximo y el Mínimo [29], tratando de conseguir principalmente dos resultados adicionales, i) el factor o variable que es más importante para el efecto y ii) el valor de tal factor [30]. Para el análisis de los perfiles además de la información estadística que los define, son valiosos los Mapas SOM monovariantes ya que permiten según la distribución de valores en el mismo, evaluar relaciones y correlaciones entre variables.

Para cumplir con las recomendaciones de la *American Statistical Association* [31], para cada variable y perfil, además de la significación estadística, se calcula su tamaño del efecto [32].

b) Fase de modelado 2: Modelo de predicción. Para la construcción del modelo que permita predecir la vulnerabilidad, se realiza un árbol de decisión basado en reglas, identificando mediante la representación de condiciones sucesivas, el grado de probabilidad de existir el patrón de vulnerabilidad obtenido en la fase de modelado 1. Para ello se ha realizado la partición de los datos en 70/30 (entrenamiento/prueba); y se han usado los árboles de inferencia condicional, basado en el test de permutación [33], utilizando pruebas no paramétricas como criterios de división de las ramas, no precisando poda de las mismas. Para ello se ha usado el paquete “rpart” del software estadístico R-Project [34], usando como parámetros: división mínima=20, profundidad máxima=2, y cubo mínimo=7.

4) Representaciones visuales: Una de las cualidades que presentan las cartografías SOM es la capacidad de representación de la información resultante de un modo relativamente sencillo de comprender, al mostrar una representación bidimensional de las instancias de partida, con la característica de que cada una de ellas tiene por “vecina” la instancia con cualidades más semejantes. En la misma cartografía se suele representar las agrupaciones de las instancias en los distintos perfiles conformados. Esta representación se suele completar con un mapa por cada uno de los atributos o variables que construyeron el mapa SOM. Finalmente se representarán espacialmente tales perfiles mediante GIS y se obtendrán los gráficos de los árboles de decisión.

RESULTADOS

Según la metodología se realizan las tareas propias de los apartados 1) y 2), obteniéndose dos bases de datos independientes. Se puede observar una síntesis descriptiva de tales variables de partida en las 3 primeras columnas de la Tabla 1 para los datos principalmente demográficos y de la

Tabla 1. Variables de las dimensiones demográfica, social, laboral, de los equipamientos y de los servicios. Se muestra el Perfil (4) con rasgos de vulnerabilidad social. conf: Confianza p-valor: ns: p>0.05; *: p<=0.05; ** p<=0.01; ***: p<=0.001. TE:Tamaño del Efecto. +++: Grande positivo; ++:Mediano positivo; +:Pequeño positivo; ---: Grande negativo; --:Mediano negativo; -:Pequeño negativo.

| N de instancias (%) | Muestra completa | | Perfil 4 | | | |
|------------------------------------|------------------|--------|----------|--------|--------|-----|
| | 5381 | 100.00 | 550 | | 10.22% | |
| Atributo | Media | DS | Media | DS | conf | TE |
| 2.1_EquipSaludPor1000Hab | 1.39 | 6.16 | 1.56 | 2.45 | ns | |
| 2.1_EquipEducacionPor1000Hab | 1.14 | 2.36 | 2.16 | 3.20 | *** | + |
| 2.1_EquipBienestarPor1000Hab | 0.83 | 1.68 | 1.85 | 3.77 | *** | ++ |
| 2.1_EquipCulDepPor1000Hab | 0.72 | 1.47 | 1.94 | 3.07 | *** | +++ |
| 2.1_EquipamientosPor1000Hab | 4.08 | 8.31 | 7.51 | 7.25 | *** | + |
| 2.2_%NoAguaCorriente | 0.91 | 2.84 | 3.16 | 6.47 | *** | ++ |
| 2.2_%Gas | 22.77 | 33.76 | 12.31 | 28.29 | *** | - |
| 2.2_%Telefono | 86.96 | 17.48 | 63.45 | 23.69 | *** | --- |
| 2.5_%PocaLimpiezaCalles | 35.10 | 19.82 | 21.23 | 22.00 | *** | --- |
| 2.6_%Delincuencia | 25.32 | 23.67 | 4.94 | 10.62 | *** | --- |
| 3.1_Poblacion | 1367 | 518.00 | 900.20 | 410.50 | *** | --- |
| 3.1_EdadMediaPoblacion | 38.01 | 4.41 | 42.74 | 4.02 | *** | +++ |
| 3.1_%Nacimientos | 11.32 | 3.21 | 9.54 | 2.78 | *** | --- |
| 3.2_PersonasPorEdificio | 14.46 | 22.91 | 2.43 | 3.79 | *** | --- |
| 3.2_%Hogares1Adulto | 18.80 | 7.55 | 27.87 | 7.37 | *** | +++ |
| 3.2_%Hogares1adultoYMenor | 1.83 | 1.12 | 1.28 | 0.92 | *** | - |
| 3.2_%Hogares2adulto | 41.24 | 6.75 | 41.01 | 5.11 | ns | |
| 3.2_%Hogares3adulto | 18.41 | 3.36 | 16.53 | 3.13 | *** | --- |
| 3.2_%Hogares4adulto | 19.71 | 6.64 | 13.32 | 4.70 | *** | --- |
| 3.2_Hogares | 449.20 | 167.50 | 340.20 | 145.70 | *** | --- |
| 3.2_HabitantesPorHogar | 3.04 | 0.36 | 2.62 | 0.31 | *** | --- |
| 3.2_RatioEdifViviendasPorHogar | 0.74 | 0.56 | 1.42 | 0.48 | *** | +++ |
| 3.3_%PoblacionArraigada | 80.11 | 9.49 | 82.19 | 5.00 | *** | + |
| 3.3_%PoblacionInmigranteProvincial | 3.87 | 5.81 | 3.83 | 3.52 | ns | |
| 3.3_%PoblacionInmigranteRegional | 1.35 | 1.28 | 0.97 | 1.09 | *** | - |
| 3.3_%PoblacionInmigranteNacional | 1.58 | 1.33 | 1.89 | 1.59 | *** | + |
| 3.3_%PoblacionInmigranteExtranjero | 1.32 | 2.79 | 1.13 | 1.95 | * | |
| 3.3_%España | 97.92 | 4.58 | 98.58 | 2.68 | *** | |
| 3.3_%EuropaUE | 0.77 | 2.97 | 0.82 | 2.21 | ns | |
| 3.3_%EuropaNoUE | 0.18 | 0.56 | 0.12 | 0.38 | *** | |
| 3.3_%AmericaDelNorte | 0.05 | 0.18 | 0.02 | 0.08 | *** | |
| 3.3_%AmericaCentral | 0.04 | 0.09 | 0.01 | 0.05 | *** | - |
| 3.3_%AmericaDelSur | 0.41 | 0.83 | 0.20 | 0.47 | *** | - |
| 3.3_%Asia | 0.08 | 0.34 | 0.02 | 0.12 | *** | |
| 3.3_%Africa | 0.54 | 2.01 | 0.23 | 1.05 | *** | |
| 3.3_%Oceania | 0.00 | 0.02 | 0.00 | 0.01 | *** | |
| 3.3_%Apatridas | 0.00 | 0.01 | 0.00 | 0.00 | *** | |
| 3.4_%TrabajaEnProvincia | 6.43 | 5.48 | 6.81 | 4.43 | * | |
| 3.4_%TrabajaEnRegion | 0.87 | 1.00 | 0.98 | 0.95 | ** | |
| 3.4_%TrabajaEnEspaña | 0.47 | 0.57 | 1.00 | 1.26 | *** | +++ |
| 3.4_%TrabajaOtroPais | 0.12 | 0.29 | 0.16 | 0.53 | ns | |
| 4.1_%Ocupados | 33.19 | 6.51 | 27.38 | 7.11 | *** | --- |
| 4.1_%Parados | 10.48 | 5.11 | 12.95 | 7.53 | *** | + |
| 4.1_%Inactivos | 38.02 | 7.24 | 43.94 | 7.25 | *** | +++ |
| 4.2_LocalComercialPor1000Hab | 26.90 | 98.90 | 19.40 | 20.90 | *** | |
| 4.2_OficinaYServiciosPor1000Hab | 10.70 | 44.60 | 8.10 | 12.20 | *** | |
| 4.2_Industrial1000Hab | 3.07 | 12.11 | 3.63 | 7.94 | ns | |
| 4.2_LocalAgrario1000Hab | 0.65 | 6.71 | 3.14 | 19.84 | ** | + |
| 4.2_LocalInactivoPor1000Hab | 13.37 | 20.22 | 19.63 | 29.47 | *** | + |
| 4.3_EstadoConstruccionesD | 0.40 | 0.23 | 0.50 | 0.21 | *** | + |
| 4.4_IDH_COMBINADO- | 0.00 | 0.00 | 0.00 | 0.00 | *** | + |
| 4.4_IDH_COMBINADO+ | 0.00 | 0.00 | 0.00 | 0.00 | ns | |
| 4.4_IDH_COMBINADO | 0.00 | 0.00 | 0.00 | 0.00 | *** | + |
| 4.5_%AgriculturaGanaderia | 4.16 | 6.06 | 7.28 | 5.74 | *** | ++ |
| 4.5_%Pesca | 0.14 | 0.59 | 0.03 | 0.12 | *** | |
| 4.5_%Industria | 3.81 | 2.46 | 2.81 | 2.18 | *** | - |
| 4.5_%Construccion | 4.47 | 2.44 | 4.35 | 2.08 | ns | |
| 4.5_%Servicios | 20.61 | 8.12 | 12.91 | 4.83 | *** | --- |
| 4.5_%Paro | 24.09 | 11.26 | 31.61 | 16.40 | *** | ++ |
| 4.6_%ViviendaEnPropiedad | 82.28 | 13.30 | 82.73 | 10.10 | ns | |
| 4.6_%ViviendasEnAlquiler | 9.39 | 11.29 | 5.05 | 5.16 | *** | - |

Tabla 2. Variables de la dimensión residencia. Se muestra el Perfil (4) con rasgos de vulnerabilidad social. Téngase en cuenta que esta información no ha intervenido en la constitución del perfil. conf: Confianza p-valor: ns: p>0.05; *: p<=0.05; ** p<=0.01; ***: p<=0.001. TE:Tamaño del Efecto. +++: Grande positivo; ++:Mediano positivo; +:Pequeño positivo; ---: Grande negativo; --:Mediano negativo; -:Pequeño negativo.

| N de instancias (%) | Muestra completa | | Perfil 4 | | | |
|-----------------------------------|------------------|--------|----------|--------|--------|-----|
| | 5381 | 100.00 | 550 | | 10.22% | |
| Atributo | Media | DS | Media | DS | conf | TE |
| 2.3_%Garaje | 19.03 | 18.79 | 16.70 | 14.07 | *** | |
| 2.4_%NoAccesible | 72.81 | 30.40 | 87.94 | 22.67 | *** | + |
| 4.3_EdadMediaConstruccionesM | 1963 | 20.00 | 1943 | 28.10 | *** | --- |
| 4.3_EdadMediaConstruccionesD | 23.38 | 14.79 | 37.01 | 14.78 | *** | +++ |
| 4.3_EstadoConstruccionesM | 1.16 | 0.21 | 1.23 | 0.20 | *** | + |
| 5.1_%RuidosExteriores | 32.53 | 18.35 | 13.15 | 15.83 | *** | --- |
| 5.1_%Contaminacion | 19.47 | 14.83 | 8.11 | 11.66 | *** | --- |
| 6.1_AlturaMediaConstruccionesM | 2.68 | 1.60 | 1.70 | 0.51 | *** | --- |
| 6.1_AlturaMediaConstruccionesD | 5.46 | 5.27 | 1.94 | 1.34 | *** | --- |
| 6.2_EdifViviendas | 327.50 | 292.70 | 460.10 | 225.30 | *** | + |
| 6.2_%Unifamiliares | 61.10 | 33.88 | 87.48 | 13.05 | *** | ++ |
| 6.2_%UnifamiliaresAgrupadas | 21.66 | 21.86 | 8.52 | 8.70 | *** | --- |
| 6.2_%UnifamiliaresConLocales | 17.04 | 24.09 | 3.94 | 6.38 | *** | --- |
| 6.2_%Plurifamiliares | 0.20 | 0.80 | 0.06 | 0.20 | *** | |
| 6.2_%EdifLocalesConAlgunaVivienda | 0.80 | 2.59 | 0.24 | 0.74 | *** | - |
| 6.2_%EdifLocales | 8.93 | 19.87 | 6.95 | 7.79 | *** | |
| 6.2_%EdifAlojamientos | 0.08 | 1.05 | 0.22 | 1.79 | ns | |
| 6.3_%MalasComunicaciones | 13.95 | 15.95 | 15.06 | 20.53 | ns | |
| 6.4_%PocasZonasVerdes | 48.85 | 25.97 | 37.60 | 31.06 | *** | - |
| 6.5_%FaltaAseosEnViviendas | 1.32 | 2.95 | 2.20 | 5.11 | *** | + |

Tabla 2 para la dimensión residencial. Siguiendo con el apartado 3) de la metodología, se lleva a cabo la Fase de modelado 1 usando para ello 63 variables de la dimensión demográfica, social, laboral, de los equipamientos y de los servicios, obteniéndose los perfiles que caracterizan la realidad demográfica y parte de las dimensiones de vulnerabilidad social.

Se puede observar lo siguiente en el Perfil 4 (Tabla 1): El análisis estadístico desvela que este perfil poblacional presenta una elevada edad media, gran cantidad de hogares con un único ocupante, abundancia de viviendas vacías y con ciertos problemas como carencia de agua corriente en proporción mayor que el resto. Curiosamente la estadística delata que habitan en asentamientos con buenos ratios de equipamientos culturales y de bienestar por población, probablemente derivado del bajo número de habitantes de tales poblaciones y una aceptable distribución de tales funciones. Espacialmente se observa que corresponden con los emplazamientos rurales más aislados y a mayor distancia de las principales ciudades. Comparando este perfil con el Perfil 2, se observa que coincide con una población rural más envejecida, que en muchas ocasiones vive sola, en entornos urbanos con poca población, con poca ocupación de las viviendas y con altos índices de analfabetismo, paro e inactividad. Mediante la representación GIS (Fig. 1), se localiza en este perfil entornos de la Hoya de Baza (Granada), los Campos de Tabernas (Almería), altos de la Sierra de Gádor (Almería), la Sierra de Aracena (Huelva), etc.

A continuación en la Fase de modelado 2 se obtiene el árbol que permite “predecir” cómo identificar el Perfil 4, es decir, se trata de identificar las cualidades residenciales del perfil poblacional con muestras de vulnerabilidad social. En

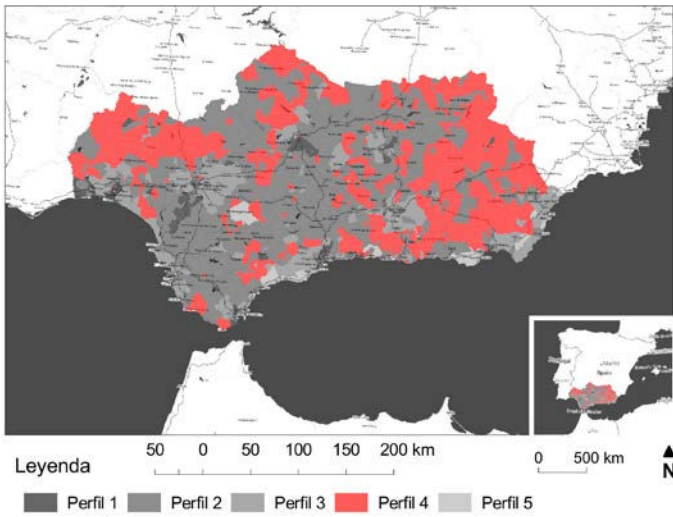


Fig. 1. Representación SIG de la clasificación generada mediante SOM. Se observa el Perfil 4 que coincide con los rasgos de vulnerabilidad.

realidad mediante los árboles de decisión se obtiene la probabilidad de identificar correctamente el Perfil 4 a partir de un número de variables muy reducido (Fig. 2). Finalmente se evalúa el rendimiento del modelo de árbol condicional de clasificación mediante la curva ROC (Fig. 3), obteniéndose un AUC=0.78.

DISCUSIÓN Y CONCLUSIONES

A partir del análisis de las experiencias bibliográficas de aplicación del modelo de clasificación y conocimiento mediante la metodología SOM y corroborado mediante la propia experiencia llevada a cabo, se puede concluir que la metodología SOM es útil para realizar un análisis exploratorio [28], con clasificaciones más potentes y robustas que las técnicas descriptivas tradicionales [35], alcanzándose una comprensión de tales patrones y su distribución espacial [29], explorándose visualmente y validándose eficazmente gracias a las consistentes propiedades geométricas de los resultados de los SOM [36], [37] y permitiendo finalmente identificar, mediante el análisis del tamaño del efecto, un perfil en el

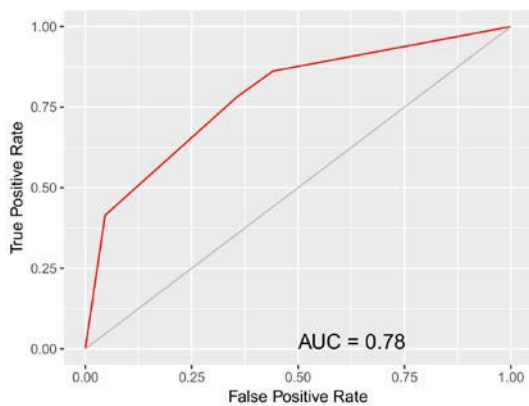


Fig. 3. Curva ROC de rendimiento del modelo Árbol de decisión condicional de pertenencia al Perfil 4 de vulnerabilidad social. El Modelo obtiene un área bajo curva (AUC) de 0.78.

territorio andaluz con clara asociación con la vulnerabilidad social. Esta asociación tiene como la utilidad permitir la formación de un DSS y en lo que venimos a llamar “Inteligencia Urbana” en cuanto que facilita afrontar mejores decisiones en la planificación, el diseño y la política urbana.

Por su lado, la metodología basada en árboles de decisión a partir de una clasificación SOM se considera útil para: atribuir de forma sencilla patrones de comportamiento que pueden ser extremadamente complejos; y predecir de forma eficaz comportamientos de variables que presentan cierto coste o dificultad de evaluación, como son las variables demográficas o sociales, a partir de otras variables con menor complejidad y coste de evaluación, como las variables residenciales.

Sin embargo, es necesario tener presente ciertas precauciones y limitaciones en el uso de estas metodologías y en su implementación concreta. Entre ellas destacan que los datos usados pueden encontrarse ya obsoletos, y que no han quedado representadas absolutamente todas las dimensiones de la vulnerabilidad social [5]. Por otro lado un análisis de la población de una sección censal no es propiamente un análisis de la población, y debe extremarse la precaución y limitar la inferencia a la escala de la observación, no alcanzando directamente a los individuos [28], es decir no se debe extrapolar a individuos las conclusiones obtenidas del estudio de grupos de individuos.

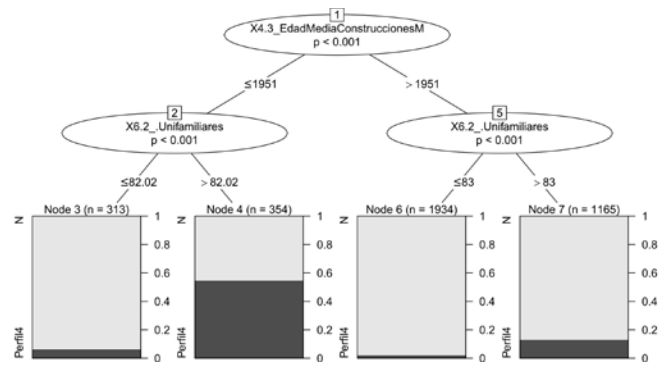


Fig. 2. Árbol de decisión. La variable que se predice es la pertenencia al Perfil 4 que se ha conectado con los rasgos de vulnerabilidad social.

Mediante la investigación aplicada al caso de estudio de la región de Andalucía, se ha obtenido un árbol de decisión de predicción del modelo obtenido mediante una metodología de clasificación no supervisada con Mapas auto-organizados, técnicas que han demostrado ser sencillas de usar, a la vez que útiles y capaces de predecir, con un relativo bajo error, fenómenos demográficos complejos y relevantes, como es la vulnerabilidad social. Para tal predicción se utiliza exclusivamente información de la realidad residencial.

En el caso específico de estudio, se han detectado gráficamente las áreas de Andalucía en las que se pronostica que pudieran estar afectadas por vulnerabilidad social, evidenciándose lo que podría nombrarse como “otra Andalucía”, una Andalucía eminentemente rural, con signos de aislamiento frente a las oportunidades de empleabilidad, etc. que ofrecen las ciudades. Las áreas urbanas enmarcadas en el

perfil de vulnerabilidad social son ciertamente escasas. Esto podría ser una debilidad del modelo, siendo recomendable un ajuste del mismo que modifique el umbral de vulnerabilidad y que englobe así ámbitos que el estado del arte identifica como tales. Por su lado el árbol de decisión obtenido resulta interesante en cuanto que permite, de forma sencilla y con cierto nivel de precisión, predecir la probabilidad de que los habitantes de un área son vulnerables sociales, utilizando para ello un escaso número de variables. Específicamente en la región de estudio, se observa que únicamente con la edad de las edificaciones y la cantidad de vivienda unifamiliar del lugar en estudio es posible predecir la pertenencia a un perfil urbano afín a situaciones de vulnerabilidad social, con una probabilidad que se puede evaluar con el indicador AUC=0.78.

Se puede por lo tanto concluir que existe una conexión y relación entre los fenómenos demográficos y de vulnerabilidad y la configuración residencial de Andalucía, debiendo ser cauto y evitar a priori un establecimiento causa-efecto entre tales fenómenos, que requerirían otras pruebas diferenciadas que distan de ser el objetivo de esta investigación.

AGRADECIMIENTOS

Esta investigación ha sido parcialmente financiada por el Plan Propio de Investigación 2016 de la Universidad de Granada PP2016-PIP09.

REFERENCIAS

- [1] S. L. Cutter, «Vulnerability to hazards», *Prog. Hum. Geogr.*, vol. 20, n.º 4, pp. 529-539, 1996.
- [2] B. Wisner, P. Blaikie, T. Cannon, y I. Davis, *At Risk: Natural Hazards, People's Vulnerability, and Disasters*. New York: Routledge, 2004.
- [3] A. Ebert, N. Kerle, y A. Stein, «Urban social vulnerability assessment with physical proxies and spatial metrics derived from air- and spaceborne imagery and GIS data», *Nat. Hazards*, vol. 48, n.º 2, pp. 275-294, 2009.
- [4] M. Prowse, *Towards a Clearer Understanding of 'Vulnerability' in Relation to Chronic Poverty*, n.º 24, 2003.
- [5] S. L. Cutter, B. J. Boruff, y W. L. Shirley, «Social vulnerability to environmental hazards», *Soc. Sci. Q.*, vol. 84, n.º 2, pp. 242-261, 2003.
- [6] D. Sánchez-González y C. Egea-Jiménez, «Enfoque de vulnerabilidad social para investigar las desventajas socioambientales. Su aplicación en el estudio de los adultos mayores», *Papeles de población*, vol. 17, n.º 69, pp. 151-185, 2011.
- [7] P. Blaikie, T. Cannon, I. Davis, y B. Wisner, «Vulnerabilidad. El Entorno Social, Político y Económico de los Desastres», *Prim. Edición Julio 1996*, p. 292, 1996.
- [8] CEPAL-ELAC, «Vulnerabilidad Sociodemográfica: viejos y nuevos riesgos para comunidades, hogares y personas.», Brasilia, 2002.
- [9] H. Gu, S. Du, B. Liao, J. Wen, C. Wang, R. Chen, y B. Chen, «A hierarchical pattern of urban social vulnerability in Shanghai, China and its implications for risk management», *Sustain. Cities Soc.*, vol. 41, n.º March, pp. 170-179, 2018.
- [10] E. Tate, «Social vulnerability indices: A comparative assessment using uncertainty and sensitivity analysis», *Nat. Hazards*, vol. 63, n.º 2, pp. 325-347, 2012.
- [11] Y. N. Maharani, S. Lee, y S. J. Ki, «Social vulnerability at a local level around the Merapi volcano», *Int. J. Disaster Risk Reduct.*, vol. 20, n.º July, pp. 63-77, 2016.
- [12] Y. N. Maharani y S. Lee, «Assessment of social vulnerability to natural hazards in South Korea: case study for typhoon hazard», *Spat. Inf. Res.*, vol. 25, n.º 1, pp. 99-116, 2017.
- [13] S. Jarupathirun y F. Zahedi, «GIS as Spatial Decision Support Systems», en *Geographic information systems in business*, J. B. Pick, Ed. Idea Group Pub, 2005.
- [14] A. G.-O. Yeh, «Urban planning and GIS», *Geogr. Inf. Syst. Princ. Tech. Manag. Appl.*, 2005.
- [15] P. G. W. Keen, «Decision support systems: The next decade», *Decis. Support Syst.*, vol. 3, n.º 3, pp. 253-265, sep. 1987.
- [16] C. Yang, R. Raskin, M. Goodchild, y M. Gahegan, «Geospatial Cyberinfrastructure: Past, present and future», *Comput. Environ. Urban Syst.*, vol. 34, n.º 4, pp. 264-277, 2010.
- [17] G. D. Buzai, G. Cacace, L. Humacata, y S. L. Lanzelotti, *Teoría y métodos de la geografía cuantitativa: libro 1: por una geografía de lo real*, n.º 1, 2015.
- [18] P. Agarwal y A. Skupin, *Self-Organising Maps: Applications in Geographic Information Science*. John Wiley & Sons, Ltd, 2008.
- [19] D. J. Power, R. Sharda, y F. Burstein, «Decision Support Systems», en *Wiley Encyclopedia of Management*, C. L. Cooper, Ed. Chichester, UK: John Wiley & Sons, Ltd, 2015, pp. 1-4.
- [20] F. J. Abarca-Alvarez, F. S. Campos-Sánchez, y R. Reinoso-Bellido, «Methodology of Decision Support through GIS and Artificial Intelligence: Implementation for Demographic Characterization of Andalusia based on Dwelling», *Estoa*, vol. 6, n.º 11, pp. 33-51, 2017.
- [21] T. Kohonen, «The Self-Organizing Map», en *Proceeding of the IEEE*, 1990, vol. 78, pp. 1464-1480.
- [22] T. Kohonen, «The self-organizing map», *Neurocomputing*, vol. 21, n.º 1-3, pp. 1-6, 1998.
- [23] H. Ritter y T. Kohonen, «Self-organizing semantic maps», *Biol. Cybern.*, vol. 61, n.º 4, pp. 241-254, 1989.
- [24] S. Kaski y T. Kohonen, «Exploratory Data Analysis By The Self-Organizing Map: Structures Of Welfare And Poverty In The World (1996)», *Neural Networks Financ. Eng. Proc. Third Int. Conf. Neural Networks Cap. Mark.*, pp. 498-507, 1996.
- [25] I. H. Witten, E. Frank, y M. a Hall, *Data Mining: Practical Machine Learning Tools and Techniques (Google eBook)*. 2011.
- [26] M. S. Silver, «On the Design Features of Decision Support Systems: The Role of System Restrictiveness and Decisional Guidance», en *Handbook on Decision Support Systems 2: Variations*, F. Burstein y C. W. Holsapple, Eds. Springer-Verlag Berlin Heidelberg, 2008, pp. 261-291.
- [27] P. Demartines y F. Blayo, «Kohonen Self-Organizing Maps: Is the Normalization Necessary?», *Complex Syst.*, vol. 6, n.º 2, pp. 105-123, 1992.
- [28] S. E. Spielmans y J.-C. Thill, «Social area analysis, data mining, and GIS», *Comput. Environ. Urban Syst.*, vol. 32, n.º 2, pp. 110-122, 2008.
- [29] L. Faggiano, D. de Zwart, E. García-Berthou, S. Lek, y M. Gevrey, «Patterning ecological risk of pesticide contamination at the river basin scale», *Sci. Total Environ.*, vol. 408, n.º 11, pp. 2319-2326, 2010.
- [30] P. K. Wu y T. C. Hsiao, «Factor Knowledge Mining Using the Techniques of AI Neural Networks and Self-Organizing Map», *Int. J. Distrib. Sens. Networks*, vol. 2015, 2015.
- [31] R. L. Wasserstein y N. A. Lazar, «The ASA's statement on p-values: context, process, and purpose», *Am. Stat.*, vol. 1305, n.º April, pp. 00-00, 2016.
- [32] R. Coe y C. Merino, «Magnitud del efecto: Una guía para investigadores y usuarios», *Rev. Psicol.*, vol. 21, n.º 1, pp. 147-177, 2003.
- [33] H. Strasser y C. Weber, «On the Asymptotic Theory of Permutation Statistics», *Math. Methods Stat.*, vol. 8, pp. 220-250, 1999.
- [34] T. Hothorn, K. Hornik, y A. Zeileis, «Unbiased recursive partitioning: A conditional inference framework», *J. Comput. Graph. Stat.*, vol. 15, n.º July, pp. 651-674, 2006.
- [35] R. Hamaina, T. Leduc, y G. Moreau, «Towards Urban Fabrics Characterization based on Buildings Footprints», en *Bridging the Geographic Information Sciences*, J. Gensel, Ed. 2012, pp. 231-248.
- [36] F. J. Abarca-Alvarez y F. Osuna-Pérez, «Cartografías semánticas mediante redes neuronales: los mapas auto-organizados (SOM) como representación de patrones y campos», *EGA. Rev. expresión gráfica Arquít.*, vol. 18, n.º 22, 2013.
- [37] J. Yan y J.-C. Thill, «Visual data mining in spatial interaction analysis with self-organizing maps», *Environ. Plan. B Plan. Des.*, vol. 36, n.º 3, pp. 466-486, 2009.