

Aplicaciones de Inteligencia Artificial al análisis de la sociedad

Autores

Prof. Ramiro Ingrassia, Facultad de Ciencias Económicas y Estadísticas, Universidad Nacional de Rosario (UNR), ringrassia@fcecon.unr.edu.ar.

Aux. Juliana Natalí Rivero, Facultad de Ciencias Económicas y Estadísticas, Universidad Nacional de Rosario (UNR), juliana.n.rivero@gmail.com

Resumen

La Encuesta de Opinión en Responsabilidad Social -2011- (EORS) tiene como finalidad documentar la opinión de los vecinos de la ciudad de San Lorenzo sobre los efectos de la actividad comercial de las empresas radicadas en dicha ciudad. Para luego, confrontar tales opiniones con las certificaciones en Responsabilidad Social Empresaria (RSE) que poseen tales empresas. También, en la encuesta anterior se previó evaluar la infraestructura del barrio para dar soporte a las políticas públicas del Municipio de San Lorenzo. Por tales motivos, se necesita el análisis de datos con herramientas que permitan “explicitar” los patrones de comportamiento de este entramado social. En el presente trabajo se pretende seleccionar los atributos que permitan describir a las familias según las distintas opiniones que han formulado. De acuerdo con este objetivo general, se realizaron diferentes pruebas para seleccionar los algoritmos que mejor rendimiento tengan en la clasificación en función de aquellos que reflejan la opinión. Aunque previamente, se probaron distintos evaluadores de atributos que permitan la reducción del número de variables relevadas por la EORS. Y de esta manera, se podría eliminar o reducir el “ruido” que de otra forma se tendría con el análisis de la totalidad de las variables. Posteriormente, se exponen los resultados que muestran en función de la opinión emitida todas las características familiares y socio-laborales.

1. Problemática

En esta nueva etapa, en base a la información recopilada con anterioridad, se buscará dilucidar las relaciones entre la opinión y las características de las personas que integran las unidades familiares relevadas en los barrios de la ciudad de San Lorenzo.

Conocer esto que no sabemos permitirá determinar el entramado social de los ciudadanos de cada barrio y sus actitudes frente a la localización de diferentes empresas. Con el concepto “entramado social” nos referimos a los niveles y características de la colaboración entre los ciudadanos. Hemos podido vislumbrar que quien participa de campañas solidarias generalmente es aquella persona que tiene problemas de salud en su familia, mientras que los que no los sufren muestran un menor grado de participación en estas actividades comunitarias. Esta situación se evidencia al comparar a los vecinos de los barrios linderos a la costa, lugar donde residen las empresas portuarias más grandes, y los que se encuentran más retirados de las empresas y que por lo tanto poseen otras características

2. Trabajos anteriores

En el marco del proyecto de investigación: “Aplicación de herramientas informáticas de conocimiento a la gestión de responsabilidad social”, se han realizado diferentes actividades con el objetivo de analizar las externalidades negativas y positivas que generan las empresas de la ciudad de San Lorenzo.

En ‘Indicadores para la Gestión del Conocimiento. Experiencias Convergentes’, se propuso el diseño de un repositorio de datos en donde se produzca una convergencia entre la opinión de los vecinos de las empresas, la información sobre responsabilidad social empresaria, información sobre aspectos sanitarios, indicadores económicos e información de los centros educativos.

Luego en ‘Uso de herramientas tecnológicas para la Gestión del Conocimiento. El caso RSE en San Lorenzo’, se realizó una Encuesta de Opinión a los barrios más interesados de la ciudad en donde se documentó la opinión sobre la Responsabilidad Social (RS) de las empresas y se relevaron los aspectos socio-económicos de las familias sanlorencinas.

Posteriormente (‘Análisis de las empresas en San Lorenzo para la selección de indicadores de responsabilidad social empresaria. Planteamiento de hipótesis sobre su estructura organizacional’ y ‘Herramientas de inteligencia de negocios aplicadas a la Responsabilidad Social. Caso San Lorenzo’), se inició el diseño de una base de datos o repositorio que permitiera almacenar los atributos más importantes de las empresas sanlorencinas que estaban disponibles en fuentes secundarias. Para proceder a la estandarización de las actividades de las organizaciones se seleccionó la nomenclatura establecida en el Código Industrial Internacional Unificado (CIIU). A partir del padrón de asociados de las cámaras empresariales de San Lorenzo (la Cámara de Comercio, Industria y Servicios de San Lorenzo y su zona (CCIS), y la Unión de Comerciantes e Industriales del Departamento San Lorenzo (UCI)), se obtuvo el rubro; el domicilio para su localización geográfica y el barrio circundante. De esta manera, se pudo realizar

un mapa geográfico que permitiera visualizar la localización de las distintas actividades económicas en la ciudad, dado que se les asignó un color en función de la letra del CIU a la cual pertenecen. También, se procedió a buscar en la red, a partir de la constancia de CUIT de cada empresa, con el objeto de verificar la clasificación realizada anteriormente y con ello se pudo incorporar la siguiente información: domicilio fiscal, teléfono, actividad principal, actividades secundarias y año de alta impositivo.

A su vez, se actualizó el relevamiento de los sitios web y por consiguiente, su calificación en el nivel de incorporación o adopción del comercio electrónico. Esto fue posible ya que en las páginas web de cada organización se pueden visualizar las certificaciones obtenidas en relación a la producción y la responsabilidad social empresarial; como también, si declara la misión empresarial y en caso afirmativo, si incorpora la preocupación por la comunidad y/o medio ambiente que la rodea.

3. Planteamiento de Objetivos

Como objetivo general se persigue establecer relaciones entre las características propias de los Jefes de Hogar y las opiniones que formularon sobre el grado de interacción con sus vecinos.

De acuerdo con dicho objetivo se plantea específicamente:

- 1) Seleccionar el algoritmo que permita reducir el número de atributos en el análisis de la opinión de los Jefes de Hogar,
- 2) Seleccionar el algoritmo que permita realizar la clasificación de los atributos que contextualizan a las familias en función de las variables de opinión,
- 3) Dilucidar las relaciones entre los anteriores

4. Metodología

Materiales y Métodos (¿Cómo se estudia el problema?)

- i. Tipo de estudio: *observacional, transversal*.
 - ii. Área de estudio: *barrios de la ciudad de San Lorenzo*.
 - iii. Métodos y recolección de datos:
Fuentes primarias: *Base de datos con la opinión de familias de San Lorenzo*.
 - iv. Tabulación y análisis de los datos:
-

En relación a los materiales y métodos, se llevará a cabo un tipo de estudio observacional y transversal, en el cual se observarán principalmente los barrios sanlorenceños situados en la costa de la ciudad.

Para abordar la problemática se estima necesaria la utilización de la información recabada en la Encuesta de Opinión de Responsabilidad Social -2011-, destinada a los Jefes de Hogar de las familias que habitan los barrios que circundan a las

organizaciones, donde se recopiló la opinión sobre Responsabilidad Social Empresaria, la situación socioeconómica de los mismos y las organizaciones en que trabajan.

La encuesta de opinión realizada y que se utilizará como fuente de datos, está estructurada en seis bloques. El Bloque A: “Descripción de la unidad familiar”, está destinado a relevar la cantidad de personas por edad y sexo que integran la familia, así como también su situación laboral y de salud. El Bloque B: “La unidad familiar y su entorno urbano”, se centra en conocer la cercanía de la vivienda a los establecimientos industriales. Por otra parte el Bloque C: “Descripción socio-laboral”, vuelve a abordar el tema de la situación laboral, integrándolo con aspectos tales como el nivel de educación alcanzado, medio de transporte utilizado para trasladarse al lugar de trabajo, la rama de actividad a la que se dedica la empresa donde trabaja y el tipo de capacitación recibido en la misma.

El Bloque D “Interacción con los vecinos del barrio”, es sobre el cual se centra esta etapa de la investigación y aborda temas relativos al grado de unión entre los vecinos y a la participación de los mismos en actividades comunitarias o solidarias. El Bloque E: “Aspectos como consumidor”, se centra en determinar la opinión de los ciudadanos sobre la contaminación y el medio ambiente, así como también el grado de participación que cada uno tiene con respecto al tema.

Por último, el Bloque F “Niveles de infraestructura del barrio”, recaba información sobre las condiciones ambientales en las que viven los vecinos en relación a servicios de cloaca, gas, agua y alumbrado entre otros.

Este tipo de análisis se puede enmarcar como señalan Campanaro e Ingrassia (2012), “dentro de las *Estrategias Generales* expuestas por Bryman y Burgess (Scribano, 2008), en la “*Teoría Generativa*” o “*Estrategia de comparación constante*”. La cual se describe en los siguientes pasos:

- 1) Consiste en descubrir la teoría desde los datos.
- 2) Después de algunas recolecciones y puesta en relación con el problema, se procede a construir categorías.
- 3) Luego, se analiza su saturación, esto asegura su significado e importancia,
- 4) Posteriormente, se establecen relaciones y conexiones entre las categorías, proponiendo algunas hipótesis sobre el problema.
- 5) Luego, se vuelve al campo y se inicia un análisis de los esquemas generados. Finalmente, las interconexiones encontradas entre esquemas teóricos y datos vuelven a ser testadas. (Scribano, 2008, p. 139)

Y dentro del segundo enfoque, en que puede abordarse el análisis de datos, el denominado *Procesos Generales*, se adopta la estrategia de “*Generación de Conceptos*”. En el cual las etapas son:

1. Inmersión en los datos y luego búsqueda de patrones,
2. Identificar posibles fenómenos sorpresas y,
3. Ser sensitivo hacia las inconsistencias.”

5. Resultados

5.1. Selección de Algoritmos

La cantidad total de variables (atributos) relevados en la EORS es de 74 y se dividen en numéricos y categóricos. La opinión de los jefes de hogares se conforma por: 12 atributos sobre la interacción con sus vecinos y 9 atributos más sobre su papel como consumidores.

En esta etapa del proyecto se analizaron distintos algoritmos disponibles en WEKA con el objeto de optimizar el rendimiento de la clasificación para la obtención de la asociación de variables. Para ejemplificar parte del proceso de selección, tomamos la variable '*grado de integración con sus vecinos*' cuyas características no difieren del resto de los atributos bajo análisis.

En este marco, se seleccionaron los siguientes algoritmos para proceder a la evaluación de atributos:

1. weka.attributeSelection.ChiSquaredAttributeEval,
2. weka.attributeSelection.GainRatioAttributeEval,
3. weka.attributeSelection.OneRAttributeEval
4. weka.attributeSelection.ReliefFAttributeEval,
5. weka.attributeSelection.SymmetricalUncertAttributeEval
6. weka.attributeSelection.InfoGainAttributeEval

El método de búsqueda elegido fue *weka.attributeSelection.Ranker*, que establece un orden de mérito y permite encontrar aquellos atributos que mejor se asocian entre sí. Adicionalmente, se plantea la comparación del evaluador de atributos "*Chi cuadrado*" –dado que los atributos son variables categóricas- contra el rendimiento de los demás evaluadores.

Además, se utilizó el método de validación cruzada (*k-fold cross validation*) para obtener una mejor precisión en el modelaje dado que el 'conjunto de validación' y 'prueba de conjunto' pertenecen a la misma población.

En primer lugar, se realizó un filtrado con los evaluadores de atributos antes mencionados con el objetivo de mejorar las asociaciones entre las variables en estudio. Como resultado se obtiene el siguiente orden (ranking) en los atributos, de los cuales se exponen los diez primeros. En los primeros 4 puestos todos los evaluadores coinciden en los atributos relativos a los barrios (cloaca, calle, alumbrado y nombre del barrio); en la 5ª y 6ª posición sólo tres evaluadores coinciden (*OneR*, *ChiSquare* e *InfoGain*) en aquellos referidos al nivel de estudio de los jefes de hogar; y en las restantes posiciones sólo coinciden *ChiSquare* e *InfoGain*. Aunque existe una similitud con *Symmetrical Uncert* dado que tales atributos representan una descripción de la situación del Jefe de Hogar y el reclamo a una empresa. En los restantes evaluadores prevalecen los atributos de opinión y algunos de la conformación del núcleo familiar. En el caso del evaluador *ReliefF* pondera más los atributos de opinión, los cuales son categóricos y los valores que adoptan son cuatro: Mucho, Bastante, Poco y Nada. En el caso de *OneR* y *GainRatio* continúan con atributos propios del Jefe de Hogar (categóricos pero con más valores) y algunas de opinión (categóricas pero con cuatro

valores).

En igual forma, el grado de interacción ('Interacción'): el evaluador *InfoGain* la coloca en el puesto 40; *SymmetricalUncert* la posiciona en el lugar 42; *OneR* en la coloca en 44 escaño; *ChiSquare* en el 43 posición; *GainRatio* la coloca en la 42 posición; y por último, *ReliefF* en la posición 34 de un total de 74 atributos.

Tabla 1: Los 10 primeros atributos según cada uno de los 'Evaluadores de Atributos'

Symmetrical Uncertainty	ReliefF	OneR	GainRatio	ChiSquared	InfoGain
Cloaca	Cloaca	Barrio	Cloaca	Barrio	Barrio
Barrio	Barrio	Cloaca	Barrio	Cloaca	Cloaca
Calle	Calle	Calle	Calle	Calle	Alumbrado
Alumbrado	Alumbrado	Alumbrado	Alumbrado	Alumbrado	Calle
Estudios_Actual	Información	Estudios Actual	Agua	Estudios Inicial	Rama Actividad
Reclamo_Empresa	Situación Laboral	Estudios Inicial	Horas/Año	Estudios Actual	Estudios Actual
Estudios_Inicial	Enfermed Ambiente	Fem_61yMas	Reclamo Empresa	Rama Actividad	Estudios Inicial
Categoría	Ayudar ambiental	Reducción CO2	Jefe_Edad	Categoría	Categoría
Rama_Actividad	Residuos	Horas/Año	Subsidios	Reclamo Empresa	Reclamo Empresa
Horas/año	Derechos	Situación Laboral	Fem_61 yMas	Jefe_Edad	Sindicato

Fuente: Elaboración Propia

La mejor evaluación por parte de *ReliefF* debe de ser que, primitivamente, fue un algoritmo desarrollado para trabajar con características lógicas y posteriormente se amplia para procesar más clases o características (Azofra, 2006). De esta manera, pondera mejor los atributos categóricos con dos o cuatro valores (características) como: Cloaca, Calle y Alumbrado, en el primer caso. Luego, coloca las variables numéricas al final de la jerarquía.

Por otro lado, los evaluadores *Symmetrical Uncertainty*, *GainRatio*, *ChiSquare* e *InfoGain* tienen resultados similares después de la cuarta posición dado que trabajan con medidas internas. Estos evaluadores permiten la medición de la **entropía o la información mutua de dos variables**, es decir, si hay asociación entre dos variables entonces conociendo una de ellas se puede saber sobre la otra (Martínez, 2011). En consecuencia, pueden trabajar con atributos que contengan una amplia variedad de clases (valores que asumen) y por lo tanto, resulta posible el análisis que involucre atributos propios del jefe de hogar y de los reclamos a las empresas.

El evaluador *OneR* según Mahmood (2009):

Un algoritmo-R tiene un conjunto de instancias como entrada, cada uno con varios atributos y una clase. Se infiere una regla como la salida del conjunto dado de ejemplos. Esta norma se basa en un solo atributo que es el más preciso en la predicción de la clase dada [valor que asume la variable] (Nevill-Manning, et al, 1995), en otras palabras, se genera un árbol de decisión de un nivel.

La idea básica del algoritmo Una-R es a elegir un atributo con la tasa de error más baja. Se toma una primera atribuir, y para cada valor de este atributo, se asigna a la más frecuente de clase que los rendimientos de este valor. Entonces el número de los errores de este atributo valor corresponde a una diferente clase. La suma de errores para todos los valores de un atributo particular forma su error total. El atributo que tiene el menor de todos errores totales se elige como la regla Uno-R. (p. 2)

Por tal motivo, este último junto con *GainRatio* puede trabajar dentro de los diez primeros puestos con variables numéricas que discretizan en el procesamiento.

Después de la construcción de estos conjuntos jerarquizados por un orden de mérito medio se procedió a la clasificación en función de la variable '**grado de interacción con sus vecinos**'.

Tabla 2: Resultados de la Clasificación de la Clase 'Interacción' según la filtración efectuada

Parámetros	Evaluadores de Atributos						
	<i>ChiSquare</i>	<i>OneR</i>	<i>ReliefF</i>	<i>Gain Ratio</i>	<i>Symmetrical Uncertainty</i>	<i>Info Gain</i>	<i>Sin Filtro</i>
<i>Clasificaciones correctas</i>	113	105	114	118	117	116	111
<i>Índice Kappa</i>	-0,0369	-0,0854	-0,0158	0,0132	-0,0149	0,0409	0,0012
<i>Error Medio Absoluto</i>	0,3082	0,314	0,3085	0,3047	0,3041	0,3075	0,3095
<i>Raíz del Error Cuadrático Medio</i>	0,3934	0,3991	0,3934	0,3897	0,3895	0,3906	0,3967

Fuente: Elaboración Propia

Así, se evaluó la precisión de las reglas para predecir los valores de la variable/atributo 'Interacción'. En la clasificación se utilizó el algoritmo '*DecisionTable*' y el método de búsqueda que se manejó fue el algoritmo *GeneticSearch*.

“Los algoritmos evolutivos tienen un carácter de búsqueda global que hace que sean especialmente adecuados para resolver problemas presentes en las distintas etapas del proceso de descubrimiento de conocimiento (Freitas, 2002). Por ejemplo, en procesos de extracción de reglas, los algoritmos evolutivos tratan de forma adecuada las interacciones entre atributos porque evalúan una regla como un todo mediante la función de adaptación en lugar de evaluar el impacto

de añadir y/o eliminar una condición de una regla, como ocurre en los procesos de búsqueda local incluidos en la mayoría de los algoritmos de inducción de reglas y árboles de decisión.” (Orallo, 2004, p. 389).

Como resultado, la mayor cantidad de instancias correctamente clasificadas se efectuó dentro del conjunto de atributos filtrados por el evaluador *GainRatio*. El índice Kappa tuvo los peores resultados en el conjunto de atributos evaluados por *OneR*, *ChiSquare*, *ReliefF* y *Symmetrical Uncertainty*. La concordancia de la regla fue insignificante en el resto de los conjuntos de atributos pero con un mejor resultado en aquel depurado por *InfoGain*. Y por último, con respecto a los errores medio y raíz cuadrática medio, los valores más bajos fueron para el conjunto de atributos refinado por *Symmetrical Uncertainty* seguido por aquel de *GainRatio*. En cuanto al promedio ponderado de la curva ROC siempre estuvo por debajo de 0,5 lo cual indica que la clasificación corresponde al azar.

A partir de este análisis, el conjunto de atributos filtrado por *Symmetrical Uncertainty* (menor error) obtiene el primer lugar seguido de aquel “depurado” por *GainRatio* (mejor clasificación) aunque la diferencia es mínima. Aunque, el nivel de concordancia entre las observaciones y las estimaciones, entre estos evaluadores, es mejor en *GainRatio*; y en *InfoGain* dicho índice tiene el mejor desempeño de todos.

En cuanto a la variable ‘Interacción’, esta no es una variable que permita una buena clasificación de los atributos relevados. Esto podría deberse a que es una variable de opinión o dependiente y las variables que contextualizan a los opinadores son mejores dado que serían las variables “independientes”.

5.2. Resultados según el filtrado por *GainRatio*

Con base a la mejor clasificación de registros u observaciones por parte de *GainRatio* expondremos a continuación los resultados obtenidos por dicha filtración de atributos. En tal sentido, si bien no se puede predecir, sí se puede **describir** quienes en dicha encuesta respondieron: Mucho, Bastante, Poco o Nada, con cierto grado de ‘eficiencia’ entre las variables.

Con el conjunto de atributos filtrado por *GainRatio* y la clasificación realizada por ‘*TableDecision*’/*GeneticSearch*, el valor predicho fue “Bastante”. La tabla resultante está conformada por: Barrio, Jefe_Edad, Rama_Actividad, Sindicato, Proximidad, Situacion_Laboral, Reclamo_Motivo, Asociacion, Conocerse, Bici_TUP, Beneficio_social, Campañas, Impactos_Nocivos, Conoce_Incidencia, Capacitación, Conoce_asesor, EnBarrio, Actividades, Cáncer, Gas.

Se puede apreciar que de aquellos atributos tendientes a describir el ‘Grupo Familiar’ solo el atributo “Cáncer” se lo utiliza para describir la variable estudiada: **nivel de interacción con los vecinos**.

No obstante, sí se selecciona un conjunto mayor de atributos que describen la ‘Unidad Familiar y su entorno urbano’: Proximidad, Conoce incidencia, Conoce asesor y Reclamo Motivo.

De igual forma, la ‘Descripción socio-laboral’ también aparece representada por: Jefe Edad, Situación Laboral, Sindicato, EnBarrio (si el trabajo está dentro del Barrio donde

vive) y Capacitación. La opinión sobre otras temáticas tiene cierta asociación como: Conocerse, Bici TUP, Beneficio social, Campañas, Impactos nocivos, Actividades y Asociación. Por último, con respecto a la infraestructura del barrio sólo es relevante el tipo de conexión a Gas que se tenga.

Para ilustrar los resultados obtenidos anteriormente, se expondrá la descripción correspondiente de aquellos que responden que el ‘Grado de interacción con sus vecinos’ ***mucho***. De esta manera se presentarán los resultados según las variables del ‘Grupo Familiar’, ‘Unidad Familiar y su entorno urbano’, ‘Descripción socio-laboral’ y demás opiniones.

a) Descripción del Grupo Familiar:

Tabla 3: Descripción de aquellos que respondieron ‘Mucho’ en Interacción según Barrio y Cáncer

Barrio	No Padece/Padeció	Padece/Padeció	Total
2 de Abril	15%		15%
Del Combate	15%		15%
Díaz Velez	5%		5%
Mitre		5%	5%
Morando	5%		5%
Rivadavia	5%		5%
San Martín	5%		5%
Villa Felisa	40%	5%	45%
Total	90%	10%	100%

Fuente: Elaboración Propia

En la tabla anterior se visualiza la distribución de los jefes de hogar que respondieron que el grado de interacción con los vecinos es ‘***Mucho***’ según el barrio donde habita y si en el seno de su familia algún familiar padece o padeció cáncer.

Debe de aclararse que la variable/atributo ‘barrio’ no pertenece a ninguna sección específica pero su ubicación en este pareció oportuno a la hora del presente estudio.

b) La Unidad Familiar y su entorno urbano:

Tabla 4: Descripción de aquellos que respondieron ‘Mucho’ en Interacción según Proximidad,

Reclamo Motivo, Conoce Incidencia y Conoce Asesor

Proximidad Reclamo Motivo	Conoce Incidencia	No	Si		Total
	Conoce Asesor	No Corresponde	No	Si	
A menos de 100m	No corresponde	5%	5%	0	10%
A más de 100m	Contaminación Edilicia	0	0	5%	5%
	Contaminación en general	0	5%	0	5%
	Enfermedad Bronquial	0	0	5%	5%
	No corresponde	10%	5%	10%	5
A más de 500m	No corresponde	15%	10%	5%	6
	No declara	0	5%	0	5%
No tiene industria en su barrio	Contaminación en general	0	0	5%	5%
	No corresponde	10%	0	0	10%
Total		40%	30%	30%	100%

Fuente: Elaboración Propia

Por ejemplo, el 5% de los encuestados tienen un ‘establecimiento industrial’ a menos de 100m (Proximidad) y no tienen ningún motivo para reclamar (Reclamo Motivo). A su vez, no conocen casos de incidencia de empresas regionales en la salud de la población (Conoce Incidencia), como tampoco dónde asesorarse si conociera alguno (Conoce Asesor). Por el contrario, el 60% de los encuestados sí conoce casos de incidencia de empresas en la salud pero la mitad no conoce donde asesorarse sobre aspectos legales y/o sanitarios. Aunque la mayoría no efectuó ningún reclamo y no tiene motivo (‘no corresponde’).

Por otra parte, otro 5% tiene un ‘establecimiento industrial’ a más de 100m (pero no más de 500m), conocen de casos de la incidencia de las empresas regionales en la salud de población por ‘enfermedades bronquiales’ y conocen asesoramiento sobre aspectos legales y/o sanitarios. En la misma proporción, pero distinto motivo de reclamo, existen familias que reclamaron por ‘contaminación edilicia’ lo cual significa que denuncian infraestructuras que son perjudiciales para los barrios.

Tabla 5: Descripción de aquellos que respondieron 'Mucho' en Interacción según Situación Laboral, Sindicato, Edad Jefe y En el barrio.

	Edad Jefe	En el Barrio															Total	
		Si	No corresponde										No en el barrio					
Situación Laboral	Sindicato	63	34	43	46	53	65	67	69	71	85	21	28	34	40	41	45	
Desocupado	No corresponde			5%	5%													10%
Monotributista	No corresponde					5%	5%											10%
Otro	No corresponde		10%	5%		5%		5%	5%	5%	5%							40%
Relación Dependencia	No contesta	5%											5%					10%
	S.C.CyE.T.C. Sta.Fe													5%				5%
	S.E.C.														5%			5%
	S.O.E.A.														5%	5%		10%
	U.O.C.R.A.											5%					5%	10%
Total		5%	10%	10%	5%	10%	5%	5%	10%	5%	5%	100%						

Fuente: Elaboración Propia

c) Descripción socio-laboral:

En relación con los atributos que exponen la ‘Descripción socio-laboral’ de los Jefes de Hogar, el 40% corresponde a empleados y el resto a cuentapropistas (Situación Laboral). En cuanto a los primeros se indagó el sindicato (Sindicato) y la localización de su trabajo (En el Barrio). Solamente el 5% respondió que su lugar de trabajo está en el barrio y su edad es de 63 años (Edad Jefe) pero no declara su sindicato.

El resto de los Jefes de Familia cuya situación laboral es la relación de dependencia no trabaja en su barrio (45%) y su edad es entre 21 y 45 años. En contraste, la mayor cantidad de cuentapropistas o desocupados tienen una edad entre 43 y 85 años. Esto permite visualizar el solapamiento de edades que dividen difusamente la situación laboral de los Jefes de Familia. Por una cuestión de dimensionalidad no se puede exponer la variable ‘Capacitación’ en una misma tabla.

d) Interacción con los vecinos del barrio y Aspectos como consumidor:

Por último, en cuanto a los atributos o variables que manifiestan la opinión de los jefes de hogar se seleccionó según el ranking establecido: Asociación, Conocerse, Bici_TUP, Beneficio social.

Tabla 6: Descripción de aquellos que respondieron ‘Mucho’ en Interacción según BiciTUP, Beneficio Social, Conocerse y Asociación

Bici TUP	Beneficio Social	Conocerse	Bastante			Mucho		Nada	Total	
		Asociación	Mucho	Nada	Poco	Mucho	Nada	Nada		
Bastante	Nada			15%	5%		10%	5%	35%	
	Poco				5%		5%		10%	
Mucho	Bastante						5%		5%	
	Nada			10%			20%		30%	
Poco	Nada						10%		10%	
	Poco		5%			5%			10%	
Total				5%	25%	10%	5%	50%	5%	100%

Fuente: Elaboración Propia

Los Jefes de Hogar que evalúan el grado de interacción con sus vecinos como ‘**Mucho**’ desde el punto de vista de sus otras opiniones se los identifica como:

- a) Un 5% de los anteriores opinan que *bastante* “La bicicleta y el transporte público son los medios más adecuados de movilidad” (BiciTUP); para *nada* “Conocen algún programa de beneficio social aplicado en la zona” (Beneficio Social); estiman que es *bastante* importante “Conocerse entre los vecinos” (Conocerse); no obstante su participación en ONG o Asociaciones sin fines de lucro (Asociación) es *poca*.

- b) Por el contrario en la misma proporción, dichos Jefes de Hogar opinan que *mucho* “La bicicleta y el transporte público son los medios más adecuados de movilidad”; *bastante* “Conocen algún programa de beneficio social aplicado en la zona”; estiman que es *mucho o muy* importante “Conocerse entre los vecinos”; no obstante su participación en ONG o Asociaciones sin fines de lucro es *nada*.

De acuerdo a lo expresado anteriormente, se describe a partir de todos los tópicos quienes son los Jefes de Hogar que opinan que *el nivel de interacción con sus vecinos es mucho*.

También debe señalarse que las demás variables que reflejan las opiniones de interacción están en el mismo nivel o por debajo del ranking establecido en el filtrado de atributos definido por *GainRatio*. Por lo tanto, el comportamiento de dichas variables es similar al descripto en el presente trabajo.

6. Conclusión y Debate

Si bien los algoritmos analizados tuvieron un rendimiento similar, los “filtradores” *GainRatio* y *Symmetrical Uncertainty* obtuvieron mejores resultados.

Para la clasificación de los atributos en función de la variable “ejemplo” Interacción se seleccionó el algoritmo *DecisionTable* con el método de búsqueda *GeneticSearch*. La “salida” de la clasificación por medio de este algoritmo permite visualizar los resultados en una tabla de múltiples entradas. Es decir, cada atributo o variable que “clasifica” la variable bajo análisis la explica de tal manera que se obtiene un cubo multidimensional.

Si bien la eficiencia de la variable no permite clasificar el resto de los atributos, se pudo describir a través de la referida tabla o cubo las demás características de los Jefes de Hogar según su respuesta.

Referencia Bibliográfica

- Azofra, A. (2006). *Un sistema inteligente para selección de características en clasificación*. (Tesis de Doctorado). Universidad de Granada. Granada.
- Campanaro, R., Ingrassia, R. (2012, noviembre). Propuesta de Datamart para el análisis de las consecuencias de la dinámica empresarial. Trabajo presentado en las 17ª Jornadas “Investigaciones en la Facultad” de Ciencias Económicas y Estadísticas, Rosario.
- Fayyad, U. (Ed.). (1996). *Advances in knowledge Discovery and Data Mining*. Menlo Park: American Association for Artificial Intelligence.
- López, R. (2007: Febrero). Análisis de los determinantes del riesgo de crédito en presencia de carteras de bajo incumplimiento. Una nueva propuesta de aplicación. *Revista Europea de Dirección y Economía de la Empresa*. 16(2), p. 71-92.
- Mahmood, A., Lei, W. (2009). One-RM: An Improved One – Rule Classifier. *Bangladesh Journal of Scientific and Industrial Research*. 44(2), p. 171-180
- Martínez, S., Correa, H. (2011, enero/junio). Evaluación de técnicas para el análisis de relevancia basadas en filtros sobre imágenes radiológicas. *Revista Ingeniería Biomédica*. 5(9), p. 26-34.
- Orallo, J., Quintana, M., Ramírez, C. (2004). *Introducción a la Minería de Datos*. Madrid: PEARSON EDUCATION.
- Sampedro, J. (2012). *Estudio y aplicación de técnicas de aprendizaje automático orientadas al ámbito médico: estimación y explicación de predicciones individuales*. (Tesis de Maestría). Universidad de Madrid. Madrid.
- Scribano, A. (2008). *La investigación social cualitativa*. Buenos Aires: Prometeo Libros.