



ESJ Natural/Life/Medical Sciences

David L. La Red Martínez,
Stella M. Gerzel,
Carlos E. Aguirre,
Rubén A. Cáceres,
José A. Berecochea Galarza,
Universidad Nacional del Chaco Austral, Sáenz
Peña, Argentina

Submitted: 19 September 2020
Accepted: 5 November 2020
Published: 30 November 2020

Corresponding author:
David L. La Red Martínez

DOI: [10.19044/esj.2020.v16n33p88](https://doi.org/10.19044/esj.2020.v16n33p88)

© Copyright 2020 Martínez et al.
Distributed under Creative Commons CC-BY 4.0
OPEN ACCESS

Cite as:
Martínez D,S.Gerzel.,C.Aguirre, R. Cáceres,Galarza
J (2020). Uso de Minería de Datos Para la
Determinación de Perfiles Socioeconómicos y
Sanitarios en la UNCAus. *European Scientific
Journal, ESJ, 16 (33)*, 87.
<https://doi.org/10.19044/esj.2020.v16n33p88>

Uso de Minería de Datos Para la Determinación de Perfiles Socioeconómicos y Sanitarios en la UNCAus

Resume

El proyecto PI N° 91, aprobado por Res. N° 332/18 CS, se desarrolla con el fin de obtener información relevante que detecte variables relacionadas con los frecuentes problemas de salud en la ciudad de Presidencia Roque Sáenz Peña (Chaco, Argentina) y su área de influencia, relacionando los pacientes con su hábitat, ecología y salud. Para ello se trabaja con la información proveniente de las actividades curriculares de vinculación comunitaria que la carrera de Medicina de la UNCAus (Universidad Nacional del Chaco Austral) realiza en los distintos barrios de dicha ciudad y su zona de influencia; con dicha información se construye un almacén de datos (data warehouse) que es estudiado con técnicas de minería de datos (data mining), especialmente técnicas de agrupamiento (clusterización) y de árboles de decisión, a los efectos de conseguir los perfiles característicos relacionados con los distintos tipos de diagnósticos; en un principio se buscan modelos descriptivos de minería de datos, para pasar en el futuro a modelos predictivos, lo cual permitiría disponer de conocimiento que permitiría mejorar la toma de decisiones en cuanto a campañas de salud hacia la población de los barrios de la ciudad de P. R. Sáenz Peña.

Subject: Demographic data mining

Palabras clave: Minería De Datos,
Soporte De Decisión, Perfiles
Socioeconómicos, Perfiles Sanitarios,
Minería Descriptiva, Minería Predictiva,
Salud Humana

Use of Data Mining to Determine the Socioeconomic and Health Profiles at UNCAus

David L. La Red Martínez

Stella M. Gerzel,

Carlos E. Aguirre,

Rubén A. Cáceres,

José A. Berecochea Galarza,

Universidad Nacional del Chaco Austral, Sáenz Peña, Argentina

DOI: [10.19044/esj.2020.v16n33p88](https://doi.org/10.19044/esj.2020.v16n33p88)

Abstract

Project PI N ° 91, approved by Res. N ° 332/18 CS, was developed in order to obtain relevant information that detects variables related to frequent health problems in the city of Presidencia Roque Sáenz Peña (Chaco, Argentina) and its area of influence, relating patients to their habitat, ecology, and health. For this, we work with the information coming from the curricular activities based on the relationship with the community that the Medicine undergraduate degree program of the UNCAus (National University of Chaco Austral) carries out in the different neighborhoods of the said city and its area of influence. With this information, a data warehouse was built and studied with data mining techniques, especially clustering techniques and decision trees, in order to achieve the characteristic profiles related to the different types of diagnosis. Initially, descriptive data mining models was sought and moved on to predictive models in the future. This would result to knowledge that would lead to better decision-making regarding health campaigns towards the population of the neighborhoods of the city of P.R. Sáenz Peña.

Keywords: Data Mining, Decision Support, Socioeconomic Profiles, Health Profiles, Descriptive Mining, Predictive Mining, Human Health

Introducción

Esta investigación responde a la necesidad de estructurar y generar información, a partir de datos obtenidos de las campañas de Actividades de Vinculación Comunitaria que realiza la carrera de Medicina como actividades curriculares, las cuales se llevan a cabo en los diferentes barrios y zonas de influencia de la localidad de Presidencia Roque Sáenz Peña.

Se cuenta con gran cantidad de datos referidos a la condición socio económica y de salud de los pacientes, pero estos datos son representados como fichas médicas en papeles. En este sentido resulta importante digitalizarlos para luego aplicar técnicas de almacenes de datos, análisis multidimensional y minería de datos descriptiva y a posteriori predictiva que permitan obtener nuevos resultados. La nueva información generada a partir del manejo de datos de esta investigación permitirá encontrar y estudiar los perfiles socioeconómicos y sanitarios de los distintos barrios de la población y sus relaciones, elaborar propuestas de intervención y fundamentar acciones de mejora.

El objetivo general del proyecto es determinar, mediante las técnicas de minería de datos, los factores que definirán las variables que causan las enfermedades crónicas frecuentes de la población, relacionando al paciente con su hábitat, ecología y salud; utilizando los determinantes sociales de la salud como variables.

Los objetivos específicos son: a) diseñar el sistema para ingresar la información en una base de datos, b) establecer las variables de estudio, c) aplicar las técnicas de minería de datos, d) analizar los resultados obtenidos, e) confeccionar informes, f) publicar los resultados obtenidos.

El proyecto es de cuatro años de duración, habiéndose completado el primero de ellos, por lo cual los resultados mostrados son parciales.

Materiales y métodos

Se trabaja con los datos obtenidos en las campañas de Actividades de Vinculación Comunitaria que realiza la carrera de Medicina como actividades curriculares, las cuales se llevan a cabo en los diferentes barrios y zonas de influencia de la localidad de Presidencia Roque Sáenz Peña, Chaco, Argentina.

Estos datos se validan y cargan mediante una aplicación (programa) desarrollado al efecto en una base de datos, desde la cual son transferidos a un almacén de datos (Data Warehouse o DW), para ser analizados con técnicas de minería de datos (Data Mining o DM).

La DM es la etapa de descubrimiento en el proceso de KDD (Knowledge Discovery from Databases), es el paso consistente en el uso de algoritmos concretos que generan una enumeración de patrones a partir de los datos pre procesados (para asegurar consistencia y congruencia de estos) (Fayyad et al., 2001; Fayyad et al., 1996; Hand et al., 2000). También se la define como un conjunto de técnicas de análisis de datos que permiten extraer patrones, tendencias y regularidades para describir y comprender mejor los datos y extraer patrones y tendencias para predecir comportamientos futuros (Simon, 1997; Berson & Smith, 1997; Frawley et al., 1992; White, 2001).

Las técnicas de DM son diversas, una de las más utilizadas es la de clustering que consiste en la partición de un conjunto de individuos en subconjuntos lo más homogéneos posible, el objetivo es maximizar la similitud de los individuos del cluster y maximizar la diferencia entre clusters. El cluster demográfico es un algoritmo desarrollado por IBM, que resuelve automáticamente los problemas de definición de métricas de distancia / similitud, proporcionando criterios para definir una segmentación óptima (Grabmeier & Rudolph, 1998; Baragoïn et al., 2002; Ballard et al., 2007; Ballard et al., 2006).

La DM es también un mecanismo de explotación, consistente en la búsqueda de información valiosa en grandes volúmenes de datos. Está muy ligada a los DW ya que los mismos proporcionan la información histórica con la cual los algoritmos de minería obtienen la información necesaria para la toma de decisiones (Gutierrez, 2001; IBM Software Group, 2003).

Un DW es una colección de datos orientada a un dominio, integrada, no volátil y variante en el tiempo para ayudar a tomar decisiones (Salton, 1989; Inmon, 1992; Inmon, 1996; Berson & Smith, 1997; Trujillo et al., 2000).

Las tecnologías de DW y DM se vienen utilizando ampliamente en el ámbito de la salud. A continuación, se referencian algunos trabajos al respecto.

En (Holzinger & Jurisica, 2014) se indica la utilización de técnicas de descubrimiento de conocimiento y específicamente de DM, aplicadas a la informática biomédica.

En (Koh & Tan, 2005) y en (Tomar & Agarwal, 2013) se describe el uso de MD para la construcción de modelos predictivos aplicados al cuidado de la salud en general y a la predicción de riesgo de ataques cardíacos en particular en (Srinivas et al., 2010).

En (Mullins et al., 2006) se describe la utilización de DM y DW como soporte para aplicativos del cuidado de la salud, trabajando con datos clínicos de pacientes.

En (Soni et al., 2011) y en (Bhatla & Jyoti, 2014) de indican distintos modelos predictivos de enfermedades cardíacas usando diferentes técnicas de DM.

En (Kavakiotis et al., 2017) se describe la aplicación de DM como ayuda para la clasificación de datos médicos.

En (Easton et al., 2014) se aplican técnicas de DM para el estudio de la mortalidad debida a accidentes cerebrovasculares.

En (Colak et al., 2015) se aplican técnicas de DM para la predicción de ataques fulminantes y en (Arslan et al., 2016) para la predicción de ataques isquémicos.

En (Paramasivam et al., 2014) se presenta el uso de DM predictiva aplicada a casos de aneurismas aórticos abdominales.

En (Morais et al., 2017) se usa DM predictiva de la necesidad de maniobras de reanimación neonatales, y de cuidados intensivos en (Peixoto et al., 2017).

En (Obenshain, 2004) se utiliza DM para estudios epidemiológicos.

También se ha aplicado DM para estudios de salud en (Alinejad-Rokny et al., 2018; Kalantari et al., 2018; Falip et al., 2017; Kaur & Wasan, 2006), entre muchos otros.

El objetivo de este proyecto es explicitar en qué medida los determinantes sociales de la salud en los distintos barrios influyen en su situación sanitaria, utilizando variables relacionadas con el hábitat, la ecología y la salud, considerando como indicadores de esta las enfermedades crónicas que poseen los habitantes.

Se consideran las siguientes variables:

- Biología humana: edad, género, antecedentes biológicos y enfermedades crónicas no transmisibles.
- Medio ambiente: vivienda, agua y excretas.
- Socioeconómico: trabajo, cobertura de salud y nivel educacional alcanzado.
- Sistema de salud: embarazos, inmunizaciones y centros de salud cercanos.
- Estilo de vida: tabaquismo.
- Educación: máximo nivel educativo alcanzado.

En base a estas variables, se utilizan técnicas de clusterización, clasificación y determinación de patrones que permiten establecer perfiles y situaciones problemáticas potenciales.

Resultados

Hasta el momento se ha completado la carga y procesamiento de la información recopilada durante el año 2019 (que totaliza 1241 registros de familias), siendo el primero de los cuatro años del proyecto, por lo cual los resultados son preliminares. La distribución por barrios es la siguiente: Ginés Benítez, 27,07%; Tiro Federal, 14,10%; Santa Mónica, 39,16%; Mitre, 12,09% y Aeroclub, 7,58%.

Las características distintivas resumidas de los habitantes por viviendas, por barrio, se muestran en la Tabla 1.

Tabla 1. Características distintivas de los habitantes

Ginés Benítez	Tiro Federal	Santa Mónica	Mitre	Aeroclub
El 88% no tiene menores de 1 año, el 12% tiene 1.	El 86% no tiene menores de 1 año, el 14% tiene 1.	El 88% no tiene menores de 1 año, el 12% tiene 1.	El 88% no tiene menores de 1 año, el 11% tiene 1, el 1% 2.	El 88% no tiene menores de 1 año, el 10% tiene 1, el 1% 2.

El 60% no tiene niños de 1 a 5 años, el 30% tiene 1, el 9% 2, el 1% 3.	El 59% no tiene niños de 1 a 5 años, el 34% tiene 1, el 6% 2, el 1% 3.	El 61% no tiene niños de 1 a 5 años, el 27% tiene 1, el 7% 2, el 3% 3.	El 71% no tiene niños de 1 a 5 años, el 16% tiene 1, el 10% 2, el 2% 3.	El 51% no tiene niños de 1 a 5 años, el 30% tiene 1, el 16% 2, el 3% 3.
El 47% de las familias no tiene niños de 6 a 14 años de edad, en tanto que el 25% tiene 1 y el 17% tiene 2.	El 40% de las familias no tiene niños de 6 a 14 años de edad, en tanto que el 33% tiene 1 y el 17% tiene 2.	El 45% de las familias no tiene niños de 6 a 14 años de edad, en tanto que el 30% tiene 1 y el 15% tiene 2.	El 49% de las familias no tiene niños de 6 a 14 años de edad, en tanto que el 27% tiene 1 y el 17% tiene 2.	El 51% de las familias no tiene niños de 6 a 14 años de edad, en tanto que el 21% tiene 1 y el 17% tiene 2.
El 57% de las familias tienen una persona masculina entre 15 y 49 años de edad y el 21% no lo tiene, el 13% tiene 2.	El 59% de las familias tienen una persona masculina entre 15 y 49 años de edad y el 15% no lo tiene, el 18% tiene 2.	El 50% de las familias tienen una persona masculina entre 15 y 49 años de edad y el 21% no lo tiene, el 21% tiene 2.	El 46% de las familias tienen una persona masculina entre 15 y 49 años de edad y el 23% no lo tiene, el 23% tiene 2.	El 66% de las familias tienen una persona masculina entre 15 y 49 años de edad y el 23% no lo tiene, el 11% tiene 2.
El 70% no tiene personas de más de 50 años, el 17% tiene 1, el 10% 2.	El 67% no tiene personas de más de 50 años, el 20% tiene 1, el 13% 2.	El 57% no tiene personas de más de 50 años, el 27% tiene 1, el 15% 2.	El 55% no tiene personas de más de 50 años, el 28% tiene 1, el 15% 2.	El 75% no tiene personas de más de 50 años, el 14% tiene 1, el 11% 2.
El 87% no tiene jubilados, el 10% tiene 1, el 3% 2.	El 79% no tiene jubilados, el 15% tiene 1, el 6% 2.	El 88% no tiene jubilados, el 8% tiene 1, el 4% 2.	El 78% no tiene jubilados, el 17% tiene 1, el 5% 2.	El 95% no tiene jubilados, el 5% tiene 1.
El 57% tiene una persona femenina de 15 a 49 años, el 17% 2, el 5% 3, el 2% 4.	El 62% tiene una persona femenina de 15 a 49 años, el 14% 2, el 7% 3, 2% 4.	El 47% tiene una persona femenina de 15 a 49 años, el 19% 2, el 8% 3, 2% 4.	El 44% tiene una persona femenina de 15 a 49 años, el 19% 2, el 7% 3, 2% 4.	El 69% tiene una persona femenina de 15 a 49 años, el 9% 2, el 2% 3.
El 99% no tiene embarazadas menores de 17	El 99% no tiene embarazadas menores de 17	El 99% no tiene embarazadas menores de 17	El 100% no tiene embarazadas	El 99% no tiene embarazadas menores de 17

años, el 1% tiene 1.	años, el 1% tiene 1.	años, el 1% tiene 1.	menores de 17 años.	años, el 1% tiene 1.
El 99% no tiene embarazadas de 18 a 21 años, el 1% tiene 1.	El 98% no tiene embarazadas de 18 a 21 años, el 2% tiene 1.	El 99% no tiene embarazadas de 18 a 21 años, el 1% tiene 1.	El 99% no tiene embarazadas de 18 a 21 años, el 1% tiene 1.	El 97% no tiene embarazadas de 18 a 21 años, el 3% tiene 1.
El 99% no tiene embarazadas mayores de 21 años, el 1% tiene 1.	El 98% no tiene embarazadas mayores de 21 años, el 2% tiene 1.	El 99% no tiene embarazadas mayores de 21 años, el 1% tiene 1.	El 98% no tiene embarazadas mayores de 21 años, el 2% tiene 1.	El 95% no tiene embarazadas mayores de 21 años, el 5% tiene 1.
El 1% no tiene controles prenatales.	El 0% no tiene controles prenatales.	El 0% no tiene controles prenatales.	El 1% no tiene controles prenatales.	El 2% no tiene controles prenatales.
El 12% tiene 2 integrantes con controles prenatales incompletos; el 10% 1, el 3% 3.	El 12% tiene 1 integrantes con controles prenatales incompletos; el 10% 2, el 5% 4.	El 10% tiene 1 integrantes con controles prenatales incompletos; el 9% 2, el 7% 3.	El 12% tiene 1 integrantes con controles prenatales incompletos; el 11% 2, el 4% 3.	El 17% tiene 1 integrantes con controles prenatales incompletos; el 19% 2, el 6% 5.
El 2% tiene controles prenatales.	El 3% tiene controles prenatales.	El 2% tiene controles prenatales.	El 2% tiene controles prenatales.	El 6% tiene controles prenatales.
El 81% informó no tener desempleados, el 11% 1 y el 5% 2.	El 81% informó no tener desempleados, el 10% 1 y el 5% 2.	El 71% informó no tener desempleados, el 15% 1 y el 7% 2.	El 61% informó no tener desempleados, el 18% 1 y el 15% 2.	El 71% informó no tener desempleados, el 20% 1 y el 7% 2.
El 57% informó no tener ama de casa y el 41% informó tener 1.	El 59% informó no tener ama de casa y el 39% informó tener 1.	El 58% informó no tener ama de casa y el 37% informó tener 1.	El 59% informó no tener ama de casa y el 39% informó tener 1.	El 54% informó no tener ama de casa y el 45% informó tener 1.
El 76% no tiene trabajadores	El 81% no tiene trabajadores	El 79% no tiene trabajadores	El 78% no tiene trabajadores	El 82% no tiene trabajadores

por cuenta propia permanente, el 19% tiene 1, el 5% 2.	por cuenta propia permanente, el 17% tiene 1, el 2% 2.	por cuenta propia permanente, el 15% tiene 1, el 5% 2.	por cuenta propia permanente, el 13% tiene 1, el 8% 2.	por cuenta propia permanente, el 14% tiene 1, el 4% 2.
El 63% no tiene trabajadores por cuenta propia temporal, el 30% tiene 1, el 5% 2.	El 62% no tiene trabajadores por cuenta propia temporal, el 30% tiene 1, el 7% 2.	El 66% no tiene trabajadores por cuenta propia temporal, el 25% tiene 1, el 8% 2.	El 74% no tiene trabajadores por cuenta propia temporal, el 17% tiene 1, el 7% 2.	El 67% no tiene trabajadores por cuenta propia temporal, el 29% tiene 1, el 3% 2.
El 73% no tiene personas empleadas, el 18% tiene 1, el 8% 2, el 1% 3.	El 59% no tiene personas empleadas, el 29% tiene 1, el 10% 2, el 2% 3.	El 65% no tiene personas empleadas, el 26% tiene 1, el 7% 2, el 1% 3, el 1% 4.	El 65% no tiene personas empleadas, el 26% tiene 1, el 7% 2, el 2% 3.	El 61% no tiene personas empleadas, el 32% tiene 1, el 7% 2.
El 60% no tiene mutual, el 13% tiene una persona con mutual, el 7% 4, el 7% 3.	El 62% no tiene mutual, el 13% tiene una persona con mutual, el 7% 4, el 7% 2.	El 67% no tiene mutual, el 14% tiene una persona con mutual, el 7% 2, el 3% 3.	El 55% no tiene mutual, el 15% tiene una persona con mutual, el 9% 2, el 8% 3.	El 55% no tiene mutual, el 21% tiene una persona con mutual, el 7% 2, el 5% 3.
El 31% tiene 1 persona con educación primaria, el 30% 0, el 26% 2.	El 28% tiene 1 persona con educación primaria, el 26% 0, el 29% 2.	El 30% tiene 1 persona con educación primaria, el 21% 0, el 29% 2.	El 34% tiene 1 persona con educación primaria, el 25% 0, el 20% 2, el 10% 3.	El 36% tiene 1 persona con educación primaria, el 35% 0, el 19% 2, el 6% 3.
El 54% no tiene personas con educación secundaria, el 25% tiene 1, el 14% 2, el 4% 3.	El 49% no tiene personas con educación secundaria, el 33% tiene 1, el 14% 2, el 4% 3.	El 53% no tiene personas con educación secundaria, el 25% tiene 1, el 15% 2, el 3% 3.	El 48% no tiene personas con educación secundaria, el 31% tiene 1, el 13% 2, el 5% 3.	El 56% no tiene personas con educación secundaria, el 24% tiene 1, el 16% 2, el 2% 3.
El 91% no tiene personas con educación superior, el 7%	El 88% no tiene personas con educación superior, el	El 95% no tiene personas con educación superior, el 4%	El 89% no tiene personas con educación superior, el 8%	El 94% no tiene personas con educación superior, el 6% tiene 1.

tiene 1, el 2% 2.	10% tiene 1, el 1% 2, el 1% 3.	tiene 1, el 1% 2.	tiene 1, el 3% 2.	
El 90% tiene agua de red directa y el 8% por canilla comunitaria, el 2% de perforación.	El 82% tiene agua de red directa y el 15% por canilla comunitaria, el 3% de perforación.	El 71% tiene agua de red directa y el 24% por canilla comunitaria, el 4% de perforación.	El 84% tiene agua de red directa y el 11% por canilla comunitaria, el 5% de perforación.	El 8% tiene agua de red directa y el 82% por canilla comunitaria, el 10% de perforación.
El 78% tiene pozo ciego, el 14% cloaca y el 8% letrina.	El 70% tiene pozo ciego, el 25% cloaca y el 5% letrina.	El 57% tiene pozo ciego, el 34% cloaca y el 9% letrina.	El 92% tiene pozo ciego, el 1% cloaca y el 7% letrina.	El 72% tiene pozo ciego y el 28% letrina.
El 43% no tiene animales desparasitados, el 22% tiene 1, el 14% 2 y el 10% 3.	El 38% no tiene animales desparasitados, el 18% tiene 1, el 18% 2 y el 9% 3.	El 44% no tiene animales desparasitados, el 18% tiene 1, el 14% 2 y el 9% 3.	El 42% no tiene animales desparasitados, el 15% tiene 2, el 12% 1 y el 11% 3.	El 64% no tiene animales desparasitados, el 14% tiene 1, el 7% 3 y el 5% 4.
El 99% informó no tener animales para consumo.	El 99% informó no tener animales para consumo.	El 97% informó no tener animales para consumo.	El 94% informó no tener animales para consumo.	El 98% informó no tener animales para consumo.
El 89% no tiene enfermos de Chagas, el 10% tiene 1, el 1% 2.	El 91% no tiene enfermos de Chagas, el 9% tiene 1.	El 90% no tiene enfermos de Chagas, el 8% tiene 1, el 2% 2.	El 95% no tiene enfermos de Chagas, el 4% tiene 1, el 1% 2.	El 95% no tiene enfermos de Chagas, el 4% tiene 1, el 1% 2.
El 75% no tiene personas con tabaquismo, el 20% tiene 1, el 4% 2.	El 79% no tiene personas con tabaquismo, el 14% tiene 1, el 6% 2, el 1% 4.	El 81% no tiene personas con tabaquismo, el 15% tiene 1, el 3% 2, el 1% 3.	El 87% no tiene personas con tabaquismo, el 10% tiene 1, el 1% 2, el 1% 4, el 1% 5.	El 92% no tiene personas con tabaquismo, el 7% tiene 1, el 1% 2.
El 89% no tiene personas con diabetes, el 10% tiene 1, el 1% tiene 2.	El 87% no tiene personas con diabetes, el 10% tiene 1, el 3% tiene 2.	El 88% no tiene personas con diabetes, el 11% tiene 1, el 1% tiene 2.	El 90% no tiene personas con diabetes, el 10% tiene 1.	El 91% no tiene personas con diabetes, el 9% tiene 1.

El 79% no tiene personas con hipertensión, el 20% tiene 1, el 1% tiene 2.	El 78% no tiene personas con hipertensión, el 18% tiene 1, el 4% tiene 2.	El 82% no tiene personas con hipertensión, el 16% tiene 1, el 2% tiene 2.	El 75% no tiene personas con hipertensión, el 20% tiene 1, el 4% tiene 2.	El 85% no tiene personas con hipertensión, el 14% tiene 1, el 1% tiene 2.
El 1% tiene tuberculosis tratado.	El 0% tiene tuberculosis tratado.	El 0% tiene tuberculosis tratado.	El 1% tiene tuberculosis tratado.	El 3% tiene tuberculosis tratado.
El 96% de las viviendas son casas, el 3% locales comerciales, el 1% rancho.	El 99% de las viviendas son casas, el 1% locales comerciales, el 1% departamento.	El 90% de las viviendas son casas, el 5% rancho, el 4% locales comerciales, el 1% departamento.	El 85% de las viviendas son casas, el 4% rancho, el 9% locales comerciales, el 3% departamento.	El 86% de las viviendas son casas, el 10% rancho, el 3% locales comerciales, el 1% sitio baldío.
El 89% de las viviendas tiene 1 familia, el 9% 2, el 1% 3, el 1% 4.	El 91% de las viviendas tiene 1 familia, el 8% 2, el 1% 3.	El 86% de las viviendas tiene 1 familia, el 13% 2, el 1% 3.	El 85% de las viviendas tiene 1 familia, el 10% 2, el 5% 3.	El 95% de las viviendas tiene 1 familia, el 3% 2, el 2% 3.
El 23% tiene 4 convivientes, el 19% 5, el 18% 3, el 10% 6.	El 27% tiene 4 convivientes, el 24% 5, el 13% 3, el 10% 2.	El 20% tiene 5 convivientes, el 18% 4, el 15% 3, el 11% 6, el 9% 2.	El 19% tiene 5 convivientes, el 15% 4, el 15% 3, el 13% 6, el 12% 2.	El 29% tiene 3 convivientes, el 20% 4, el 15% 5, el 14% 2, el 6% 6.

Así como en la tabla precedente el estudio se ha hecho desde la perspectiva (dimensión de análisis) de los barrios, también se hará (para futuras publicaciones) desde las perspectivas educación primaria, educación secundaria, educación superior, embarazadas menores de 17 años, desempleados, mutual sí, mutual no, niños de 1 a 5 años, tabaquismo, hipertensión, diabetes, Chagas, tuberculosis tratado, tuberculosis no tratado, animales desparasitados, animales para consumo, total de convivientes, vivienda, excretas, agua, etc.

Comentarios y discusiones

Los principales comentarios y discusiones se resumen en la Tabla 2.

Table 2. Principales comentarios y discusiones

Ítem considerado	B° con mayor %	B° con menor %
Menores de 1 año	Tiro Federal	

Niños de 1 a 5 años	Aeroclub	
Niños de 6 a 14 años	Tiro Federal	
Masculinos entre 15 y 49 años	Tiro Federal	
Personas de más de 50 años	Mitre	
Jubilados	Mitre	
Personas femeninas de 15 a 49 años	Tiro Federal	
Embarazadas menores de 17 años		Mitre
Embarazadas de 18 a 21 años	Aeroclub	
Embarazadas mayores de 21 años	Aeroclub	
Sin controles prenatales	Aeroclub	
Controles prenatales incompletos	Aeroclub	
Con controles prenatales	Aeroclub	
Desempleados	Mitre	
Amas de casa	Aeroclub	
Trabajadores por cuenta propia permanente	Ginés Benítez	
Trabajadores por cuenta propia temporal	Tiro Federal	
Personas empleadas	Tiro Federal	
Sin mutual	Santa Mónica	
Con educación primaria	Mitre	
Sin educación secundaria	Aeroclub	
Sin educación superior	Aeroclub	
Agua de red directa	Ginés Benítez	Aeroclub
Agua por canilla comunitaria	Aeroclub	Ginés Benítez
Agua de perforación	Aeroclub	Ginés Benítez
Pozo ciego	Mitre	Santa Mónica
Cloaca	Santa Mónica	Aeroclub
Letrina	Aeroclub	Tiro Federal
Animales no desparasitados	Aeroclub	
Animales para consumo	Mitre	
Personas con Chagas	Ginés Benítez	
Personas con tabaquismo	Ginés Benítez	
Personas con diabetes	Tiro Federal	
Personas con hipertensión	Mitre	
Personas con tuberculosis tratado	Aeroclub	
Casas	Tiro Federal	
Locales comerciales	Mitre	
Rancho	Aeroclub	
Departamento	Mitre	
Sitio baldío	Aeroclub	
Viviendas con una familia	Aeroclub	
Viviendas con 3 convivientes	Aeroclub	
Viviendas con 4 convivientes	Tiro Federal	
Viviendas con 5 convivientes	Tiro Federal	
Viviendas con 6 convivientes	Mitre	

Conclusiones y líneas futuras de investigación

La metodología de trabajo propuesta ha demostrado ser de utilidad para la recolección de la información, su carga, depuración y procesamiento con técnicas de almacenes de datos y de minería de datos.

Conforme a lo planificado, se ha podido caracterizar la población y el entorno de los distintos barrios incluidos en las actividades curriculares de la carrera de Medicina de la UNCAus, como se ha mostrado en las tablas 1 y 2. Esta caracterización permite extraer las siguientes conclusiones:

- El mayor porcentaje de niños menores de 1 año corresponde al B° Tiro Federal.
- El mayor porcentaje de jubilados y personas mayores corresponde al B° Mitre.
- El menor porcentaje de embarazadas menores de 17 años corresponde al B° Mitre.
- El mayor porcentaje de embarazos sin controles prenatales o con controles prenatales incompletos corresponde al B° Aeroclub.
- El mayor porcentaje de desempleados corresponde al B° Mitre.
- Las mayores falencias en el suministro de agua potable de red se dan en el B° Aeroclub.
- Las mayores falencias en cuanto a red cloacal se dan en los B° Mitre y Aeroclub.
- El mayor porcentaje de mascotas no desparasitadas se observa en el B° Aeroclub.
- Los mayores porcentajes de mal de Chagas y de tabaquismo se detectaron en el B° Ginés Benítez.
- El mayor porcentaje de diabetes se detectó en el B° Tiro Federal.
- El mayor porcentaje de hipertensión se registró en el B° Mitre.
- El mayor porcentaje de tuberculosis corresponde al B° Aeroclub.
- El mayor porcentaje de viviendas precarias se registró en el B° Aeroclub.
- El mayor porcentaje de situaciones de hacinamiento de detectó en el B° Mitre.

Esta información característica de la población y de su realidad socioeconómica referida a varios barrios de la ciudad de Roque Sáenz Peña, que se pudo cuantificar con el estudio realizado, permitirá tomar líneas de acción específicas a las autoridades pertinentes, tales como:

- Campañas de educación sexual y planificación familiar.
- Campañas de concientización acerca de la necesidad de los controles prenatales.

- Obras de extensión de la red de agua potable y de la red cloacal.
- Planes de mejoramiento de las unidades habitacionales y reducción del hacinamiento.
- Campañas para desparasitar mascotas.
- Campañas para prevenir o tratar casos de mal de Chagas, tabaquismo, diabetes, hipertensión y tuberculosis.
- Campañas de capacitación de adultos desempleados con vistas a su inserción laboral.
- Campañas de asistencia especial a jubilados y adultos mayores.

En la Tabla 3 se indican los barrios prioritarios para las distintas acciones específicas mencionadas.

Asimismo, habiéndose demostrado la utilidad de la metodología empleada para la recolección, carga y procesamiento de la información, como así también para la obtención de conclusiones factibles de convertirse en acciones concretas de mejora para la población, se ha planificado seguir trabajando en los próximos años con la misma metodología y cubriendo otros barrios, además de realizar nuevas visitas a los ya mencionados.

Table 3. Barrios prioritarios para acciones específicas

Barrios prioritarios para acciones específicas	B° Tiro Federal	B° Aeroclub	B° Mitre	B° Ginés Benítez
Acciones específicas				
Campañas de educación sexual y planificación familiar	X			
Campañas de concientización acerca de la necesidad de los controles prenatales		X		
Obras de extensión de la red de agua potable y de la red cloacal		X		
Planes de mejoramiento de las unidades habitacionales y reducción del hacinamiento		X	X	
Campañas para desparasitar mascotas		X		
Campañas para prevenir o tratar casos de mal de Chagas y tabaquismo				X
Campañas para tratar diabetes	X			
Campañas para tratar hipertensión			X	
Campañas para tratar tuberculosis		X		
Campañas de capacitación de adultos desempleados con vistas a su inserción laboral			X	

Campanas de asistencia especial a jubilados y adultos mayores			X	
---	--	--	---	--

Agradecimientos

Este trabajo es sostenido por el Proyecto de Investigación “Determinación de perfiles socioeconómicos y sanitarios de las personas atendidas en las campañas efectuadas como actividades curriculares de la carrera de Medicina de la UNCAus en su área de influencia”, PI N° 91, aprobado por Res. N° 332/18 CS de la Universidad Nacional del Chaco Austral (Argentina). El software ISW (InfoSphere Warehouse) de gestión de DW y de DM fue obtenido de IBM a través de la Iniciativa Académica de dicha empresa.

References:

1. Alinejad-Rokny, H., Sadroddiny, E. & Scaria, V. (2018). Machine learning and data mining techniques for medical complex data analysis. ELSEVIER. Neurocomputing.
2. Arslan, A. K., Colak, C. & SarihanbaInonu, M. E. (2016). Different medical data mining approaches-based prediction of ischemic stroke. ELSEVIER. Computer Methods and Programs in Biomedicine. 130. 87–92.
3. Ballard, CH., Beaton, A., Chiou, D., Chodagam, J., Lowry, M., Perkins, A., Phillips, R. & Rollins, J. (2006). Leveraging DB2 data warehouse edition for business intelligence. En IBM International Technical Support Organization. USA, IBM Press.
4. Ballard, CH., Rollins, J., Ramos, J., Perkins, A., Hale, R., Dorneich, A., Cas Milner, E. & Chodagam, J. (2007). Dynamic warehousing: Data mining made easy. En IBM International Technical Support Organization. USA, IBM Press.
5. Baragoin, C., Chan, R., Gottschalk, H., Meyer, G., Pereira, P. & Verhees, J. (2002). Enhance your business applications. Simple integration of advanced data mining functions. En IBM International Technical Support Organization. USA, IBM Press.
6. Berson, A. & Smith, S. J. (1997). Data warehouse, data mining & OLAP. USA. Mc Graw Hill.

7. Bhatla, N. & Jyoti, K. (2012). An Analysis of Heart Disease Prediction using Different Data Mining Techniques. *International Journal of Engineering Research & Technology*. 1-8. 1-4.
8. Colak, C., Karaman, E., Gokhan TurtaybaI, M. (2015). Application of knowledge discovery process on the prediction of stroke. *ELSEVIER. Computer Methods and Programs in Biomedicine*. 119. 181–185.
9. Easton, J. F., Stephens, C. R. & Angelova, M. (2014). Risk factors and prediction of very short term versus short / intermediate term post-stroke mortality: A data mining approach. *ELSEVIER. Computers in Biology and Medicine*. 54. 199–210.
10. Falip, J., Aït-Younes, A., Blanchard, F., Delemer, B., Diallo, A. & Herbin, M. (2017). Visual instance-based recommendation system for medical data mining. *ELSEVIER. Procedia Computer Science*. 112. 1747–1754.
11. Fayyad, U., Grinstein, G. & Wierse, A. (2001). *Information Visualization in Data Mining and Knowledge Discovery*. San Francisco, USA: Morgan Kaufmann Publishers.
12. Fayyad, U., Piatetskiy-Shapiro, G., Smith, P. & Ramasasmy, U. (1996). *Advances in Knowledge Discovery and Data Mining*. USA: AAAI Press / MIT Press.
13. Frawley, W., Piatetsky-Shapiro, G. & Matheus, CH. (1992). Knowledge Discovery in Database an Overview. En Piatetsky-Shapiro, G Frawley, W. J (Ed) *Knowledge discovery in databases* (pp. 1-31). USA: The AAAI Press.
14. Grabmeier, J. & Rudolph, A. (1998). Techniques of cluster algorithms in data mining version 2.0. En IBM Deutschland Information System GmbH. GBIS (Global Business Intelligence Solutions). Germany. IBM Press.
15. Gutierrez, J. (2001). *Data Mining, Extracción de Conocimiento en Grandes Bases de Datos*. Ponencia o documento presentado en VII Congreso Argentino de Ciencias de la Computación – V Escuela Internacional de Informática, Calafate, Argentina.
16. Hand, D., Mannila, H. & Smyth, P. (2000). *Principles of Data Mining*. USA: The MIT Press. USA.
17. Holzinger, A. & Jurisica, I. (2014). *Knowledge Discovery and Data Mining in Biomedical Informatics: The Future Is in Integrative,*

- Interactive Machine Learning Solutions. Springer-Verlag. LNCS 8401. Knowledge Discovery and Data Mining in Biomedical Informatics. 1-18.
18. IBM Software Group (2003). Enterprise Data Warehousing with DB2: The 10 Terabyte TPC-H Benchmark. IBM Press. USA.
 19. Inmon, W. (1996). Building the Data Warehouse. USA: John Wiley & Sons.
 20. Inmon, W. (1992). Data Warehouse Performance. USA: John Wiley & Sons.
 21. Kalantari, A., Kamsin, A., Shamshirband, S., Gani, A., Alinejad-Rokny, H. & Chronopoulos, A.T. (2018). Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions. ELSEVIER. Neurocomputing. 276. 2–22.
 22. Kaur, H. & Wasan, S. K. (2006). Empirical Study on Applications of Data Mining Techniques in Healthcare. Science Publications. Journal of Computer Science. 2-2. 194-200.
 23. Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N. Vlahavas, I. & Chouvarda, I. (2017). Machine Learning and Data Mining Methods in Diabetes Research. ELSEVIER. Computational and Structural Biotechnology Journal. 15. 104–116.
 24. Koh, H.C. & Tan, G. (2005). Data Mining Applications in Healthcare. Journal of Healthcare Information Management. 19-2. 64-72.
 25. Morais, A., Peixoto, H., Coimbra, C., Abelha, A. & Machado, J. (2017). Predicting the need of Neonatal Resuscitation using Data Mining. ELSEVIER. Procedia Computer Science. 113. 571–576.
 26. Mullins, I. M., Siadat, M. S., Lyman, J., Scully, K., Garrett, C. T., Miller, W. G., Muller, R., Robson, B., Apte, C., Weiss, S., Rigoutsos, I., Platt, D., Cohen, S. & Knaus, W.A. (2006). Data mining and clinical data repositories: Insights from a 667,000 patient data set. ELSEVIER. Computers in Biology and Medicine. 36. 1351–1377.
 27. Obenshain, M.K. (2004). Application of Data Mining Techniques to Healthcare Data. MAT. Statistics for Hospital Epidemiology. 25-8. 690-695.
 28. Paramasivam, V., Yee, T.S., Dhillon, S. K. & Sidhu, A. S. (2014). A methodological review of data mining techniques in predictive

- medicine: An application in hemodynamic prediction for abdominal aortic aneurysm disease. ELSEVIER. *Biocybernetics and Biomedical Engineering*. 34. 139–145.
29. Peixoto, R., Ribeiro, L., Portela, F., Santos, M. F. & Rua, F. (2017). Predicting Resurgery in Intensive Care - A data Mining Approach. ELSEVIER. *Procedia Computer Science*. 113. 577–584.
 30. Salton, G. (1989). *Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer*. Addison, Wesley.
 31. Simon, A. (1997). *Data Warehouse, Data Mining and OLAP*. USA: John Wiley & Sons.
 32. Soni, J., Ansari, U. & Sharma, D. (2011). Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction. *International Journal of Computer Applications*. 17-8. 43–48.
 33. Srinivas, K., Kavihta Rani, B. & Govrdhan, A. (2010). Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks. *International Journal on Computer Science and Engineering*. 2-2. 250-255.
 34. Tomar, D. & Agarwal, S. (2013). A survey on Data Mining approaches for Healthcare. SERSC. *International Journal of Bio-Science and Bio-Technology*. 5-5. 241-266.
 35. Trujillo, J., Palomar, M. & Gomez, J. (2000). Applying Object-Oriented Conceptual Modeling Techniques to The Design of Multidimensional Data-bases and OLAP Applications. En *First International Conference On Web-Age Information Management WAIM 2000*. *Lecture Notes in Computer Science* 1846. 83-94.
 36. White, CJ. (2001). *Enterprise Analytics for the Intelligent e-Business*. USA: IBM Press.