

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

David L. La Red Martínez, Stella Maris Gerzel, Emanuel Aguirre
Rubén A. Cáceres, José A. Berecochea Galarza, Selene Vallejo

Universidad Nacional del Chaco Austral
Sáenz Peña, (3700), Argentina

laredmartinezdavid@uncaus.edu.ar, stellagerzel@uncaus.edu.ar, aguirremanuel@uncaus.edu.ar,
rubens.uncaus@gmail.com, joseberecochea@gmail.com, selenevfores@gmail.com

Resumen

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 de 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 la ciudad y zona de influencia; con dicha información se construye un almacén de datos (data warehouse) que es estudiado con técnicas de minería de datos, 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; se buscan modelos descriptivos y predictivos de minería de datos, lo cual permitiría disponer de conocimiento que permitiría mejorar la toma de decisiones en cuanto a campañas de salud hacia la población de las zonas.

Palabras Clave: minería de datos, soporte de decisión, perfiles socioeconómicos, perfiles

sanitarios, minería descriptiva, minería predictiva, salud humana.

Contexto

El desarrollo del proyecto 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, Argentina.

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 conocimientos. Los conocimientos generados a partir de los resultados de esta investigación permitirán 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.

Introducción

La Minería de Datos (del inglés Data Mining - DM), es un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias en grandes volúmenes de datos, utilizando técnicas de AI y ML. Estas técnicas permiten extraer patrones y tendencias para describir y comprender mejor los datos y predecir comportamientos futuros.

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 preprocesados [1] [2] [3].

La DM es 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 [4] [5] [6] [7].

Las técnicas de DM son diversas, una de las más utilizadas es la de clustering (o agrupamiento de datos) 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 [8] [9] [10] [11].

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 [12] [13].

Un DW es una colección de datos orientada a un dominio, integrada, no volátil y variante en el tiempo para ayudar a tomar decisiones [14].

Un DW es una colección de datos orientado a temas, integrado, no volátil, de tiempo variante, que se usa para el soporte del proceso de toma de decisiones gerenciales.

Un DW es también un conjunto de datos integrados orientados a una materia, que varían con el tiempo, y que no son transitorios, los cuales soportan el proceso de toma de decisiones de una administración [15] [16] [4] [17].

Las tecnologías de DW y DM se vienen utilizando ampliamente en el ámbito de la salud. A continuación, se referencian algunos trabajos al respecto: [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30] [31] [32] [33] [34] [35] [36], entre muchos otros.

Líneas de Investigación y Desarrollo

Determinar, mediante las técnicas de minería de datos, los factores o variables que causan las enfermedades crónicas frecuentes de las personas, relacionando el paciente con su hábitat, ecología y salud; utilizando los determinantes sociales de la salud como variables.

Utilizar un sistema desarrollado para ingresar la información en una base de datos, luego establecer las variables de estudio y aplicar las técnicas de minería de datos. Analizar los resultados obtenidos y confeccionar informes.

Resultado Esperados

Lograr 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, y utilizando técnicas de clusterización, clasificación y determinación de patrones se espera establecer perfiles y situaciones problemáticas potenciales.

Los resultados logrados hasta el momento se han publicado en [37] y [38].

Formación de Recursos Humanos

El equipo de trabajo está compuesto por un Doctor, una Licenciada en Sistemas de Información, dos Ingenieros en Sistemas, un Profesor en Ciencias Químicas y del Ambiente y una becaria estudiante avanzada de Medicina.

Referencias

- [1] U. Fayyad, G. Grinstein, A. Wierse. **Information Visualization in Data Mining and Knowledge Discovery**. San Francisco, USA: Morgan Kaufmann Publishers. 2001.
- [2] U. Fayyad, G. Piatetskiy-Shapiro, P. Smith, U. Ramasamy. **Advances in Knowledge Discovery and Data Mining**. USA: AAAI Press / MIT Press. 1996.
- [3] D. Hand, H. Mannila, P. Smyth. **Principles of Data Mining**. USA: The MIT Press. USA. 2000.
- [4] A. Simon. **Data Warehouse, Data Mining and OLAP**. USA: John Wiley & Sons. 1997.
- [5] A. Berson, S. J. Smith. **Data warehouse, data mining & OLAP**. USA. Mc Graw Hill. 1997.
- [6] W. Frawley, G. Piatetsky-Shapiro, Ch. Matheus. **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. 1992.
- [7] C. J. White. **Enterprise Analytics for the Intelligent e-Business**. USA: IBM Press. 2001.
- [8] J. Grabmeier, A. Rudolph. **Techniques of cluster algorithms in data mining version 2.0**. En IBM Deutschland Informations systeme GmbH. GBIS (Global Business Intelligence Solutions). Germany. IBM Press. 1998.
- [9] C. Baragoin, R. Chan, H. Gottschalk, G. Meyer, P. Pereira, J. Verhees. **Enhance your business applications. Simple integration of advanced data mining functions**. En IBM International Technical Support Organization. USA, IBM Press. 2002.
- [10] Ch. Ballard, J. Rollins, J. Ramos, A. Perkins, R. Hale, A. Dorneich, E. Cas Milner, J. Chodagam. **Dynamic warehousing: Data mining made easy**. En IBM International Technical Support Organization. USA, IBM Press. 2007.
- [11] Ch. Ballard, A. Beaton, D. Chiou, J. Chodagam, M. Lowry, A. Perkins, R. Phillips, J. Rollins. **Leveraging DB2 data warehouse edition for business intelligence**. En IBM International Technical Support Organization. USA, IBM Press. 2006.
- [12] J. Gutierrez. **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. 2001.
- [13] IBM Software Group. **Enterprise Data Warehousing with DB2: The 10 Terabyte TPC-H Benchmark**. IBM Press. USA. 2003.
- [14] G. Salton. **Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer**. Addison-Wesley. 1989.
- [15] W. Inmon. **Data Warehouse Performance**. USA: John Wiley & Sons. 1992.

- [16] W. Inmon. **Building the Data Warehouse**. USA: John Wiley & Sons. 1996.
- [17] J. Trujillo, M. Palomar, J. Gomez. **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. 2000.
- [18] A. Holzinger, I. Jurisica. **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. 2014.
- [19] K. Srinivas, B. Kavihta Rani, A. Govrdhan. **Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks**. International Journal on Computer Science and Engineering. 2-2. 250-255. 2010.
- [20] H. C. Koh, G. Tan. **Data Mining Applications in Healthcare**. Journal of Healthcare Information Management. 19-2. 64-72. 2005.
- [21] I. M. Mullins, M. S. Siadat, J. Lyman, K. Scully, C. T. Garrett, W. G. Miller, R. Muller, B. Robson, C. Apte, S. Weiss, I. Rigoutsos, D. Platt, S. Cohen, W. A. Knaus. **Data mining and clinical data repositories: Insights from a 667,000 patient data set**. ELSEVIER. Computers in Biology and Medicine. 36. 1351–1377. 2006.
- [22] N. Bhatla, K. Jyoti. **An Analysis of Heart Disease Prediction using Different Data Mining Techniques**. International Journal of Engineering Research & Technology. 1-8. 1-4. 2012.
- [23] J. Soni, U. Ansari, D. Sharma. **Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction**. International Journal of Computer Applications. 17-8. 43–48. 2011.
- [24] D. Tomar, S. Agarwal. **A survey on Data Mining approaches for Healthcare**. SERSC. International Journal of Bio-Science and Bio-Technology. 5-5. 241-266. 2013.
- [25] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, I. Chouvarda. **Machine Learning and Data Mining Methods in Diabetes Research**. ELSEVIER. Computational and Structural Biotechnology Journal. 15. 104–116. 2017.
- [26] J. F. Easton, C. R. Stephens, M. Angelova. **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. 2014.
- [27] C. Colak, E. Karaman, M. Gokhan TurtaybaI. **Application of knowledge discovery process on the prediction of stroke**. ELSEVIER. Computer Methods and Programs in Biomedicine. 119. 181–185. 2015.
- [28] A. K. Arslan, C. Colak, M. E. SarihanbaInonu. **Different medical data mining approaches based prediction of ischemic stroke**. ELSEVIER. Computer Methods and Programs in Biomedicine. 130. 87–92. 2016.
- [29] V. Paramasivam, T. S. Yee, S. K. Dhillon, A. S. Sidhu. **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. 2014.
- [30] H. Alinejad-Rokny, E. Sadroddiny, V. Scaria. **Machine learning and data mining techniques for medical complex data analysis**. ELSEVIER. Neurocomputing. 276. 1. 2018.
- [31] A. Kalantari, A. Kamsin, S. Shamshirband, A. Gani, H. Alinejad-Rokny, A. T. Chronopoulos. **Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions**. ELSEVIER. Neurocomputing. 276. 2–22. 2018.

- [32] J. Falip, A. Aït-Younes, F. Blanchard, B. Delemer, A. Diallo, M. Herbin. **Visual instance-based recommendation system for medical data mining**. ELSEVIER. Procedia Computer Science. 112. 1747–1754. 2017.
- [33] A. Morais, H. Peixoto, C. Coimbra, A. Abelha and J. Machado. **Predicting the need of Neonatal Resuscitation using Data Mining**. ELSEVIER. Procedia Computer Science. 113. 571–576. 2017.
- [34] R. Peixoto, L. Ribeiro, F. Portela, M. F. Santos and F. Rua. **Predicting Resurgery in Intensive Care - A data Mining Approach**. ELSEVIER. Procedia Computer Science. 113. 577–584. 2017.
- [35] M. K. Obenshain. **Application of Data Mining Techniques to Healthcare Data**. MAT. Statistics for Hospital Epidemiology. 25-8. 690-695. 2004.
- [36] H. Kaur, S. K. Wasan. **Empirical Study on Applications of Data Mining Techniques in Healthcare**. Science Publications. Journal of Computer Science. 2-2. 194-200. 2006.
- [37] D. L. la Red Martínez, S. M. Gerzel, C. E. Aguirre, R. A. Cáceres, J. A. Berecoechea Galarza. **Uso de Minería de Datos Para la Determinación de Perfiles Socioeconómicos y Sanitarios en la UNCAus**. European Scientific Journal (ESJ). Volume 16 – N° 33. pp. 88-104. ISSN N° 1857-7881. University Ss "Cyril and Methodius" Skopje, Macedonia. 2020.
- [38] D. L. la Red Martínez, S. M. Gerzel, C. E. Aguirre, R. A. Cáceres, J. A. Berecoechea Galarza, S. Vallejo. **Perfiles socioeconómicos y sanitarios de barrios de Sáenz Peña, Chaco - Ciencia, tecnología y sociedad unidas por la universidad**. Editorial Académica Española - OmniScriptum S.R.L. Publishing Group. ISBN N° 978-620-388625-2. 2022.