

# APLICACIÓN DE MINERÍA DE DATOS ESPACIAL EN EL AREA DE SALUD EN LA ZONA DE INFLUENCIA DE LA UNNOBA.

Claudia Russo<sup>1</sup>, Javier Charme<sup>1</sup>, María Rosana Piergallini<sup>1</sup>, Ma. Mercedes Guasch<sup>1</sup>, Adriana Torriggino<sup>1</sup>, Ana Smail<sup>1</sup>

Instituto de Investigación y Transferencia en Tecnología (ITT), Escuela de Tecnología, Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA)

Sarmiento y Newbery (CP 6000), Junín, Buenos Aires, Argentina. Teléfonos (0236) 4636945/44

{claudia.russo, javier.charme, rosana.piergallini, mercedes.guasch, adriana.torriggino, ana.smail }@itt.unnoba.edu.ar

## Resumen

El aumento del volumen y variedad de información que se encuentra digitalizada en base de datos de diversas organizaciones ha crecido exponencialmente en las últimas décadas.

Sobre esta información es posible aplicar técnicas que nos permiten extraer conocimiento útil desde los datos almacenados y que se engloban bajo la denominación de Knowledge Discovery in Databases (KDD), donde Minería de Datos (MD) es un paso del KDD.

Cuando en la base de datos se incorpora el dato espacial, dada la complejidad de estos tipos de datos, los objetos que se almacenan (puntos, líneas, polígonos), y su estructura de datos se dificulta la utilización de la minería de datos tradicional. La Minería de Datos Espacial (MDE) provee un gran grupo de técnicas y herramientas para la

explotación de estos datos que permiten encontrar patrones potencialmente útiles.

El número de base de datos espaciales está creciendo constantemente y la diversidad de áreas en que este tipo de bases de datos es utilizada es muy amplio: medicina, geología, química, astronomía solo por mencionar algunos de ellas.

En salud y epidemiología se valora la importancia de la componente espacial en sus investigaciones y en el diseño de estrategias diferenciadas de prevención y control. En salud pública, el uso de las tecnologías de la información ha facilitado el conocimiento de los nuevos problemas de salud como la prevención de posibles eventos futuros.

A través de esta línea de investigación se propone aplicar técnicas de MDE para la extracción de nuevos conocimientos que asistan a la toma de decisiones en Salud Pública en el área de influencia de la Universidad Nacional del Noroeste de la Provincia de Buenos Aires

(UNNOBA). Los avances en este sentido contribuirán a preservar y mejorar el nivel de la salud pública en general.

**Palabras clave:** KDD, Minería de Datos, Minería de Datos Espacial, Salud Publica

## Contexto

La línea de trabajo que aquí se presenta se inscribe en el marco del Proyecto de Investigación “Tecnologías exponenciales en contextos de realidades mixtas e interfaces avanzadas”, aprobado por la Secretaría de Investigación, Desarrollo y Transferencia de la UNNOBA, convocatoria a Subsidios de Investigación Bianuales (SIB2015).

El tema de estudio de nuestra línea de trabajo es la utilización de MDE en salud pública y epidemiología en el área de influencia de la Universidad Nacional del Noroeste de la Provincia de Buenos Aires.

El proyecto se desarrolla en el Instituto de Investigación y Transferencia en Tecnología (ITT), en la sede Pergamino dependiente de la mencionada Secretaría, y se trabaja en conjunto con la Escuela de Tecnología y el Departamento de Salud de la UNNOBA.

El equipo está constituido por docentes e investigadores que desarrollan actividades en el departamento de Informática y Tecnología y en el departamento de Salud de la UNNOBA.

## Introducción

Las necesidades de toma de decisiones requieren el descubrimiento de nuevos modelos no esperados o imposibles de descubrir manualmente a partir de un gran volumen de datos.

El KDD es el “Proceso no trivial de identificar patrones válidos, novedosos,

potencialmente útiles y comprensibles a partir de los datos” [1]. El descubrimiento de patrones válidos se logra con la MD, que a través de la aplicación de diferentes técnicas puede obtener patrones y relaciones dentro de los datos permitiendo la creación de modelos. El KDD se encarga de la preparación de los datos y la interpretación de los resultados obtenidos, los cuales dan un significado a estos patrones encontrados.

El proceso KDD se organiza en cinco fases: integración y recopilación de datos; selección, limpieza y transformación; minería de datos evaluación e interpretación y difusión. Si bien ambos términos, MD y KDD, se utilizan como sinónimos en realidad MD es una de las etapas del proceso de extracción de conocimiento. [2].

En esencia la MD es un mecanismo de exploración y análisis consistente en la búsqueda y extracción de información valiosa, patrones y reglas ocultas en grandes volúmenes de datos.

Las áreas de salud y epidemiología requieren de la aplicación de estrategias diferenciadas para la prevención y control sanitario. El interés principal es el análisis y estudio de la distribución geográfica de las enfermedades y su relación con los riesgos potenciales, a fin de generar planes tendientes a la prevención y control de los eventos de salud que se presentan en la población. El objetivo de los estudios epidemiológicos es determinar la relación espacio-persona-tiempo. Identificar los riesgos existentes y poder determinar las áreas afectadas es parte importante de la vigilancia epidemiológica [3].

La MDE es considerada una rama de la MD tradicional, pone énfasis en la extracción de conocimiento relevante inherente a la naturaleza espacial de los datos. La MDE provee los mecanismos y herramientas como respuesta a la

dificultad de resolver problemas de descubrimiento de conocimiento en bases de datos espaciales con los enfoques tradicionales de la minería de datos [4].

Se la puede definir como el proceso automático o semiautomático de seleccionar, explorar, modificar, visualizar y valorar grades volúmenes de datos espaciales con el objeto de descubrir conocimiento. Permite obtener correlaciones no evidentes y potencialmente útiles entre objetos geográficos.

La aplicación de la MDE busca resolver diversos problemas mediante el descubrimiento de conocimiento, aplicando diversas técnicas donde los objetos espaciales cuentan además con características no espaciales y sirven como entrada a algoritmos de minería.

La obtención de conocimiento a partir de Bases de Datos espaciales, generadas en el ámbito de salud, permite identificar regiones donde hay que prestar especial atención en vigilancia epidemiológica y adecuar los planes de prevención y control de eventos de salud. Los avances en este sentido contribuirán a preservar y mejorar el nivel de la salud pública en general [5].

## **Líneas de Investigación, Desarrollo e Innovación**

La línea de investigación tiene como objetivo obtener conocimiento previamente desconocido y potencialmente útil para la toma de decisiones en el área de salud tendientes a prevenir y controlar el estado de salud de la población.

A través de esta línea de investigación se propone aplicar técnicas de minería de datos espaciales para la extracción de nuevos conocimientos que asistan a la

toma de decisiones en Salud Pública en el área de influencia de la UNNOBA.

## **Resultados y Objetivos**

Dado que es una línea de reciente creación se está llevando a cabo la etapa de recopilación de trabajos realizados en similares temáticas.

Para continuar con:

- Evaluación de métodos de KDD que tienen en cuenta las características especiales de la información de una base de datos espacial.

- Relevamiento y análisis de las herramientas existentes en la actualidad de MDE. Selección de la herramienta a utilizar en la fase de MD.

- Obtención de modelos predictivos y/o descriptivos que brinde apoyo para la toma de decisiones en el área de salud.

## **Formación de Recursos Humanos**

En esta línea I/D se espera concluir con un trabajo de Práctica Profesional Supervisada de la Ingeniería en Informática y el desarrollo de dos tesis de posgrado dirigidas por miembros de este proyecto.

## **Referencias**

[1] Fayyad, U.M. et al. "From Data Mining to Knowledge Discovery: An overview. Advances in Knowledge Discovery and Data Mining, Pp.1-34, AAAI/MIT Press 1996.

[2] Orallo, J.H., Quintana, M.J., Ramirez, C.. Introducción a la Minería de Datos. Pearson Prentice Hall. 2004.

[3] La minería de datos espaciales y su aplicación en los estudios de salud y epidemiología. Gonzalez Polanco, L., Perez Bentancourt, Y.G. Revista Cubana de Inf. en Ciencias de la Salud; 24(4):482-489. 2013

[4] Trends in Spatial Data Mining. Shekhar, S. Zhang, P.,Huang, Y., Vatsavai, R.. Department of Computer Science and Engineering. University of Minesota. 4-192, 200 Union ST SE, Miniapolis, MN 55455.

[5] Guasch, M.M., Piergallini,M.R., Smail, A, Russo, C., Torrigino, A., Dorzán, M.G., .Gagliardi, E. y Taranilla, M.T.. “VisualOET Software para el análisis y seguimiento de objetos espacio-temporales aplicable a Salud Pública”. CAIS (Congreso Argentino de Informática y Salud) 44 JAIIO. 2015. Pp.38-46. ISSN: 2451-7607