

BIOINGENIERÍA INFORMÁTICA APLICADA A LA PREDICCIÓN DE ENFERMEDADES CARDIOLÓGICAS Y SU IMPLEMENTACIÓN EN EL HOSPITAL DELICIA CONCEPCIÓN MASVERNAT DE LA CIUDAD DE CONCORDIA, PROVINCIA DE ENTRE RÍOS

¹ FALAPPA, Marcelo Alejandro, ² BENEDETTO Marcelo Gabriel, ² CARABIO, Ana Lía Ramona, ² SILVA LAYES, María Elizabeth, ² BENÍTEZ, Duval Horacio

¹ Departamento de Ciencias e Ingeniería de la Computación - Universidad Nacional del Sur Avenida Alem 1253 - Bahía Blanca (B8000CPB) - Tel.: +54(0291)4595135 mfalappa@cs.uns.edu.ar

² Facultad de Ciencias de la Administración - Universidad Nacional de Entre Ríos Monseñor Tavella 1424 – Concordia, Entre Ríos (3200) - Tel.: +54(0345)4231433 {marben, anacar}@fcad.uner.edu.ar, elizabeth.silva@gmail.com, dhbenitez@gmail.com

RESÚMEN

El sector sanitario, sin lugar a dudas es uno de los ámbitos en el que se administran grandes volúmenes de datos; principalmente en el área clínica. Esto conduce a identificar una importante necesidad de encontrar maneras de administrar, integrar, analizar e interpretar ese gran conjunto de datos; procurando identificar patrones de comportamiento que sean de utilidad en la toma de decisiones médicas. En este sentido, la minería de datos es una herramienta que nos ayuda en el análisis de los mencionados datos, sirviendo de apoyo tanto en el área de cuidados de la salud como en la investigación médica; a través de su aplicación que abarca desde el diagnóstico médico hasta la capacidad para analizar datos desde una variedad de perspectivas tendientes a descubrir patrones ocultos [4].

Actualmente, existen diversas investigaciones que permiten encarar tratamientos de enfermedades específicas, aunque siguen siendo estudios puntuales que, si bien tienen muy buenos resultados, no son de aplicabilidad a la práctica cotidiana de consultas médicas. Teniendo en cuenta esta realidad, la investigación que se plantea persigue lograr el desarrollo de un componente de software que permita incorporar la predicción de probabilidad de riesgos de enfermedades cardíacas como sistema de soporte a las decisiones clínicas incorporada al acto asistencial; así como también en el proceso de gestión de medicina preventiva. Si bien existen diferentes proyectos aplicados que permiten conocer sobre la probabilidad de riesgos, los

mismos no terminan siendo integrados a la Historia Clínica Electrónica, de manera de tener una aplicación inmediata en el acto asistencial [11].

En este proyecto se plantea como principal objetivo desarrollar un componente de software capaz de generar, con aprendizaje automatizado, un modelo con capacidades predictivas sobre enfermedades cardíacas; que permitan un mejor soporte a decisiones de diagnóstico clínico y un avance significativo en la medicina preventiva para actos asistenciales de pacientes.

Palabras clave: Minería de Datos, Sistemas de Soporte a Decisiones Clínicas, Modelos Predictivos.

CONTEXTO

La línea de investigación descrita en este artículo se encuentra enmarcada en el contexto del proyecto PID 7060 denominado: “*Bioingeniería Informática aplicada a la predicción de enfermedades cardíacas y su implementación en el Hospital Delicia Concepción Masvernats de la ciudad de Concordia, Provincia de Entre Ríos*” de la Facultad de Ciencias de la Administración, Universidad Nacional de Entre Ríos. El citado proyecto se enmarca dentro del Convenio Específico de Colaboración entre Programas de Investigación y Postgrado, celebrado en el año 2008 entre la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos y el Departamento de Ciencias e Ingeniería de la Computación de la

Universidad Nacional del Sur y en convenio con el Hospital Delicia Concepción Masvernat de la ciudad de Concordia, Provincia de Entre Ríos.

Uno de los principales objetivos de este proyecto es la formación de recursos humanos para investigación en la Facultad de Ciencias de la Administración de la UNER, especializados en la línea prioritaria de investigación denominada “Ingeniería de Software y Lenguajes de Programación” establecida por Res. 25/11 del C.D. Al ser también ésta una de las principales líneas de investigación del Departamento de Ciencias e Ingeniería de la Computación de la UNS, se justifica la creación de un equipo de investigación inter-universidades que sea contenedor del desarrollo de investigadores de la U.N.E.R. en el área.

INTRODUCCIÓN

Actualmente es inconmensurable la información con la que cuentan las instituciones de salud sobre sus pacientes, tanto a nivel administrativo como asistencial. Los datos referidos al área clínica son utilizados por el médico para obtener información del paciente, ya sea para efectuar diagnósticos, tratamientos, medicación; así como también por los propios sistemas informáticos para generar múltiples alertas (como pueden ser interacciones medicamentosas, control sobre medicamentos contraindicados para una patología específica, resultados anormales en estudios, entre otros).

El análisis de este gran cúmulo de datos nos brinda una oportunidad inmejorable para [1]:

- Realización de estudios epidemiológicos (análisis de rendimientos de campañas de información, prevención, sustitución de fármacos, entre otros),
- Cálculo de expectativas de vida,
- Identificación de terapias médicas satisfactorias para diferentes enfermedades,
- Asociación de síntomas y clasificación diferencial de patologías,

- Estudio de factores de riesgo para la salud en distintas patologías,
- Segmentación de pacientes para una atención más inteligente y
- Identificación de terapias médicas y tratamientos erróneos para determinadas enfermedades.

De esta manera, aparecen nuevas oportunidades para mejorar, no sólo la medicina personalizada sino también el rendimiento y resultado de los sistemas sanitarios. Esto conlleva a identificar necesidades referidas a nuevas maneras de administrar, integrar, analizar e interpretar tales conjuntos de datos; que posibiliten encontrar patrones de comportamiento de utilidad en la toma de decisiones médicas. En este aspecto, la minería de datos es una herramienta fundamental que resulta imprescindible en el análisis de dichos datos.

Si bien, en la actualidad existen investigaciones que conllevan a desarrollos informáticos que sirven de apoyo en tratamientos de distintas enfermedades, las implementaciones que permitan incorporar el apoyo a la toma de decisiones clínicas en el marco de la atención del paciente en tiempo real son escasas.

Teniendo en cuenta que en nuestro país una de las principales causas de muerte son las enfermedades cardiovasculares, se considera necesario el desarrollo de sistemas fiables que sirvan de apoyo a las decisiones clínicas en dos momentos fundamentales (la respuesta del sistema cuando el paciente concurre a una visita médica y una respuesta proactiva que permita brindar apoyo a través de la aplicación de medicina preventiva); logrando reducir el tiempo del diagnóstico y aumentar la precisión del mismo en este tipo de enfermedades [11].

LÍNEAS DE INVESTIGACIÓN Y DESARROLLO.

La línea de trabajo principal del proyecto de investigación radica en el estudio, análisis y comparación de técnicas de minería de datos para ser aplicados en la generación de un

modelo predictivo sobre enfermedades cardiovasculares.

Se prevé desarrollar un componente de software que pueda integrarse a una historia clínica electrónica para brindar apoyo en la toma de decisiones en el diagnóstico clínico.

OBJETIVOS

Objetivo General:

Desarrollo de un componente de software capaz de generar, con aprendizaje automatizado, un modelo con capacidades predictivas sobre enfermedades cardiológicas; que permitan un mejor soporte a decisiones de diagnóstico clínico y un avance significativo en la medicina preventiva para actos asistenciales de pacientes.

Objetivos específicos:

- Evaluar diferentes técnicas de minería de datos para generar un modelo aplicado a la predicción de probabilidad de riesgo en enfermedades cardiovasculares.
- Generar un componente de software (*Web Service*) para ser integrado a la Historia Clínica Electrónica (HCE) de las instituciones sanitarias; con implementación de una prueba piloto para instancias asistenciales en el Hospital público “*Delicia Concepción Masverná*” de la ciudad de Concordia, provincia de Entre Ríos.
- Desarrollar un módulo de soporte a la medicina preventiva, con capacidades de generar alertas tempranas sobre pacientes con riesgo cardiológico, con umbrales parametrizables basado en rango etario y sexo.
- Medir las alertas generadas por el sistema y realizar una comparativa con datos estadísticos de referencia de la Organización Mundial de la Salud (OMS); de manera de evaluar las capacidades del modelo generado determinando el sesgo propio por las

características de la población de referencia.

RESULTADOS ESPERADOS

Siendo que el proyecto se encuentra en su fase inicial, durante la ejecución del mismo se pretenden alcanzar los siguientes resultados por actividades:

Relevamiento bibliográfico: Exploración de técnicas de minería de datos aplicables, detectando clasificadores relevantes para el dominio cardiológico. Selección de variables de uso frecuente a incorporar al estudio. Revisión de modelos algorítmicos aplicables.

Preprocesamiento de datos: Recopilación de datos, conformación del *dataset* de entrenamiento, depuración, extracción de ruido, validación de datos. Preparación de datos en general, codificación de variables. Categorización de datos, binarios, multivaluado, evaluación de datos no categóricos, establecimiento de rangos y acondicionamiento de tipos. Preprocesamiento de datos no estructurados.

Selección de clasificadores: Medición del rendimiento de clasificadores evaluados, descartando todos aquellos que muestran poca incidencia cardiológica, detección de nuevos clasificadores que no han sido considerados aún, medición de indicadores, clasificación de instancias. Conformación del *dataset*.

Generación de modelos predictivos: Análisis y comparación de distintos métodos de *machine learning* aplicables, medición de indicadores relevantes de evaluación del modelo, medición de sensibilidad, especificidad, tasa de falsos positivos, y tasa de falsos negativos. Evaluación de incorporación de pesos relativos de detecciones erróneas que permitan corregir el modelo, establecimiento del rango de actuación del sistema y frecuencia de actualización óptima.

Análisis de factores extras necesarios: Detección de factores, nuevos datos e indicadores; que permitan generar nuevas

estrategias de captación de datos, para una mejora del modelo, proyección de necesidades futuras, recomendación de políticas de captación de datos relevantes.

Desarrollo del componente de software.

Desarrollo del *Web Service* de interoperabilidad con historias clínicas de instituciones de salud, e implementación del piloto en una institución del medio. Recomendaciones para la adopción del modelo.

Evaluación empírica del componente

generado: Validación de alertas generadas, revisión en conjunto con profesionales médicos, análisis de las capacidades del modelo, determinación de resultados logrados, recomendaciones futuras, determinación de aportes factibles a políticas sanitarias de la población en general.

Registro, exposición del conocimiento adquirido y transferencia al medio: Registro y publicación de logros de proyecto, publicación científica, transferencia al medio, implementación de prueba piloto en el Hospital público “*Delicia Concepción Masvernat*” de la ciudad de Concordia, provincia de Entre Ríos, si los resultados logrados muestran aporte significativo y validan la hipótesis inicial.

FORMACIÓN DE RECURSOS HUMANOS

Se brindará a los integrantes del proyecto, formación en lo que se refiere a técnicas avanzadas de Minería de Datos, Análisis de Datos y Data Science.

Los integrantes del proyecto, se desempeñan en cátedras relacionadas directamente con el tema central de la investigación, por lo que este trabajo tendrá impacto directo e inmediato en la docencia.

Se procederá a dirigir becarios de investigación, así como también tesinas finales de grado, dirigidos por el director del proyecto

de investigación y/o por los docentes investigadores del mismo. Para estos casos, también se prevé la presentación a convocatorias de becas ante organismos provinciales y nacionales.

Los integrantes participarán de reuniones científicas y técnicas que permitan actualizar los conocimientos en el tema de interés. También se trabajará en la presentación de trabajos en congresos nacionales e internacionales relacionados con el área del proyecto. Estos trabajos servirán para divulgar los conocimientos obtenidos durante el trabajo de investigación.

Uno de los principales objetivos del proyecto es que el personal docente de la UNER dedicado al mismo avance y/o concluya con sus estudios de posgrado, así como también se incorporen becarios realizando investigaciones en temas afines a la temática del proyecto.

El Director del proyecto, Marcelo Alejandro FALAPPA, es Doctor en Ciencias de la Computación egresado de la Universidad Nacional del Sur (UNS) en el año 1999. Actualmente es Profesor Asociado con Dedicación Exclusiva en el Departamento de Ciencias e Ingeniería de la Computación de la UNS a cargo de la materia curricular Bases de Datos, es Investigador Adjunto del CONICET especializado en Actualización de Bases de Datos Deductivas y Teoría de Cambio, y tiene Categoría II en el Programa de Incentivos.

El codirector del proyecto, Marcelo Gabriel BENEDETTO, es Magíster en Sistemas de Información, se encuentra realizando los cursos del Doctorado en Ciencias de la Computación de la U.N.S. y posee categoría III en el Programa de Incentivos.

La integrante, Licenciada en Sistemas de Información Ana Lía Ramona CARABIO, posee categoría V en el Programa de Incentivos y se encuentra realizando la Maestría en Redes de la U.N.LP., restando la presentación de la Tesis; la integrante María Elizabeth SILVA LAYES es Magíster en Sistemas de Información de la UNER. Finalmente, el Licenciado en Sistemas Horacio Duval BENÍTEZ se encuentra realizando los

cursos de la Maestría en Sistemas de Información de la UNER y posee categoría V en el Programa de Incentivos.

BIBLIOGRAFÍA

[1] Ana Lía Carabio, Elizabeth Silva Layes, Fabian Frola, O. Bioingeniería aplicada en el diagnóstico de enfermedades. 45 JAIIO, CAIS 2016, Séptimo Congreso Argentino de Informática y Salud (2016), 240–248.

[2] Gillespie, G. There's gold in them thar' databases. *Health Data Management* 8(11) (2000), 40–52.

[3] Andrew Wilson, Lehana Thabane, A. H. Application of data mining techniques in pharmacovigilance. *British Journal of Clinical Pharmacology* 57(2) (2003), 127–134.

[4] Kuttiannan Thangavel, Palanichamy Iaganathari, P. E. Data mining approach to cervical cancer patients analysis using clustering technique. *Asian Journal of Information Technology* 5 (4) (2006), 413–417.

[5] Sellappan Palaniappan, R. A. Intelligent heart disease prediction system using data mining techniques. *IEEE/ACS International Conference on Computer Systems and Applications* (2008), 108–115.

[6] Techentin, B. Big data and graph analytics in a health care setting. http://www.graphanalysis.org/SC12/03_Techentin.pdf, noviembre 2012. Online; accedido en 2018.

[7] Ayad Ghany Ismaeel, D. Y. M. Effective data mining technique for classification cancers via mutations in gene using neural network. *International Journal of Advanced Computer Science and Applications* 7(7) (2016), 69–76.

[8] Shantakumar Patil, Y. K. Intelligent and effective heart attack prediction system using data mining and artificial neural network. *European Journal of Scientific Research* Vol.31 No.4 (2009), 642–656.

[9] Examples of Big Data Analytics In Healthcare That Can Save People. <https://www.datapine.com/blog/big-data-examples-in-healthcare/>. 2018

[10] Bernard Marr. Big Data In Healthcare: Paris Hospitals Predict Admission Rates Using Machine Learning.

<https://www.forbes.com/sites/bernardmarr/2016/12/13/big-data-in-healthcare-paris-hospitals-predict-admission-rates-using-machine-learning/#1532778779a2> (2016)

[11] Elizabeth Silva Layes, Marcelo Alejandro Falappa, O. Sistemas de soporte a las decisiones clínicas. XIII Workshop de Investigadores en Ciencias de la Computación (2011), 113-115