ISSN 1870-4069

Minería de datos centrada en el usuario para el análisis de la supervivencia y mortalidad de casos de cáncer de mama en mujeres de origen mexicano

Aldair Antonio-Aquino[1], Guillermo Molero-Castillo[2], Rafael Rojano-Cáceres[3], Alejandro Velázquez-Mena[4]

[1] Universidad Veracruzana, MSICU, Facultad de Estadística e Informática, México

[2] CONACYT, México

[3] Universidad Veracruzana, Facultad de Estadística e Informática, México

[4] Universidad Nacional Autónoma de México, Facultad de Ingeniería, México

aquinoaldair@hotmail.com, ggmoleroca@conacyt.mx, rrojano@uv.mx, mena@fi-b.unam.mx

Resumen. Actualmente existen diversos procesos que conducen el desarrollo de un proyecto de minería de datos. Sin embargo, éstos, a pesar de su variedad, no están centrados en el usuario; trayendo como consecuencia aplicaciones con limitaciones de usabilidad y accesibilidad. En este trabajo se presenta una minería de datos centrada en el usuario con base en los fundamentos de la norma ISO 9241-210:2010. Se analizó la supervivencia y mortalidad de casos de cáncer de mama en mujeres de origen mexicano. Los datos utilizados corresponden a registros clínicos del Instituto Nacional del Cáncer de los Estados Unidos. Como resultado se obtuvo una precisión del 87.4%. Además, las pruebas de usabilidad basado en heurísticas permitieron identificar mejoras de interacción entre la aplicación y los usuarios finales.

Palabras clave: Cáncer de mama, DCU, minería de datos, origen mexicano.

User-Centered Data Mining for the Analysis of Survival and Mortality of Breast Cancer Cases in Woman of Mexican Origin

Abstract. Currently, there are several processes that lead the development of a data mining project. However, these, despite their variety, are not user-centered; consequently, there are applications with limitations of usability and accessibility. This paper presents a data mining user-centered based on the fundamentals of ISO

9241-210:2010. Survival and mortality of breast cancer cases in women of Mexican origin were analyzed. The data used correspond to clinical records of the National Cancer Institute of the United States. As a result, 87.4% accuracy was obtained. In addition, usability testing based on heuristics helped to identify improvements for interaction between the application and end-users.

Keywords. Breast cancer, data mining, Mexican origin, UCD.

1. Introducción

En la actualidad, se realizan diversas actividades físicas, domésticas, laborales, de investigación, entre otras, que implican un proceso de creación y almacenamiento de nueva información. Esta información puede llegar a ser valiosa para el proceso de la toma de decisiones, esto si se le aplica los métodos y herramientas adecuadas. Una de estas herramientas es la minería de datos como un proceso para la exploración y análisis de volúmenes de datos para el descubrimiento de conocimiento útil. En este sentido, en la actualidad la minería de datos juega un papel importante en el crecimiento y éxito de las organizaciones; propiciando así la búsqueda de nuevo conocimiento manifestado en patrones de datos, tendencias, reglas, grupos, clasificaciones, entre otros.

Precisamente, como resultado del surgimiento de la minería de datos, es necesario contar con procesos que permitan planificar y guiar el desarrollo de los proyectos. Así, en la actualidad existen diversos procesos con enfoques y funcionamiento distintos. Sin embargo, estos procesos, a pesar de su amplia variedad, no están centrados en el usuario, limitando así la participación de éstos en cada una de las etapas que comprenden; trayendo como consecuencia desarrollos de minería de datos con limitaciones de usabilidad y accesibilidad [1, 2]. Por lo que, existe la necesidad de buscar nuevas formas de analizar y procesar las fuentes de datos. Precisamente, una de estas formas es a través de una minería de datos centrada en el usuario.

La importancia que tiene el usuario en proyectos de descubrimiento de conocimiento es fundamental debido a que éstos poseen diversos conocimientos, estilos cognitivos y otras habilidades mentales que mediante un proceso de análisis se puede lograr un mejor entendimiento de las necesidades, tareas y características que se requieren considerar en el proyecto [1, 3]. Asimismo, la importancia del usuario en la minería de datos no solo está en la exploración de volúmenes de datos para el descubrimiento de conocimiento, sino también en el proceso de la toma de decisiones mediante el uso de herramientas interactivas que sean fáciles de usar, aprender y recordar [4]. Además, para involucrar al usuario como parte importante en el proceso de minería de datos se debe considerar características como: a) privado, para preservar la privacidad de los usuarios; b) personal, para asegurar de que el usuario se beneficie del conocimiento encontrado; c) portátil, para asegurar de que el flujo de datos esté en todas partes y en cualquier lugar; y d) potente, para proporcionar recursos suficientes a los usuarios para el descubrimiento de conocimiento a través de los datos [5].

En este sentido, al ser los usuarios parte esencial en este tipo de proyectos, es crucial involucrarlos desde el análisis y entendimiento del proyecto hasta la validación y presentación de los resultados. Es importante señalar que una interpretación incorrecta

de las necesidades y requerimientos de los usuarios pudiera conducir al fracaso de los proyectos de minería de datos o limitar las expectativas de los usuarios [6].

En este trabajo se presentan los resultados de la investigación sobre la contribución de una minería de datos centrada en el usuario con base en los principios de la norma ISO 9241-210:2010, enfocado al análisis de la supervivencia y mortalidad de casos de cáncer de mama en mujeres de origen mexicano. La fuente de datos utilizada corresponde a registros de la base de datos del Programa de Vigilancia, Epidemiología y Resultados Finales (SEER) del Instituto Nacional del Cáncer de los Estados Unidos.

2. Procesos de minería de datos

A pesar de la amplia variedad de tareas y técnicas de minería de datos es necesario trabajar con un marco de trabajo que permita planear y guiar el desarrollo de los proyectos. Actualmente, uno de los más conocidos es el proceso de descubrimiento de conocimiento en base de datos (Knowledge Discovery in Databases o KDD), el cual consta de una serie de etapas para la generación de conocimiento y la toma de decisiones. Este proceso tiene la característica de ser iterativo e interactivo, y que se orienta a las decisiones que toma el usuario [7], sin embargo, no describe las tareas y actividades específicas que se deben realizar en cada una de sus etapas [8]. Otro de los procesos es SEMMA (Sample, Explore, Modify, Model, Assess), creado por SAS Institute (Statistical Analysis Systems), que lo define como la selección, exploración y modelado de grandes volúmenes de datos para descubrir patrones de interés [9]. Particularmente, SEMMA inicia con un análisis exploratorio de datos, ignorando el análisis y entendimiento del proyecto [10, 11]. Por otra parte, SEMMA está relacionada particularmente al uso de productos comerciales de SAS Institute.

Tabla 1. Principales características de los procesos de minería de datos.

	KDD	**CRISP-DM**	**SEMMA**	**Catalyst**	**Six Sigma**
Fases	– Integración y recopilación – Selección, limpieza y transformación – Minería de datos – Evaluación e interpretación – Difusión y uso	– Entendimiento del negocio – Entendimiento de los datos – Preparación de los datos – Modelado – Evaluación – Despliegue	– Muestreo – Exploración – Modificación – Modelado – Evaluación	– Preparación de los datos – Modelado – Refinar el modelo – Implementar el modelo – Comunicación de resultados	– Definición – Medición – Análisis – Mejora – Control
Etapas iterativas	Si	Si	No	Si	No
Elección de herramientas	Libres y comerciales	Libres y comerciales	Comerciales	Libres y comerciales	Libres y comerciales
Evaluación del resultado	Basado en los objetivos del proyecto	Basado en el modelo y los objetivos del proyecto	Basado en el modelo	Basado en los objetivos del proyecto	Basado en el modelo
Orientada a MD	Si	Si	Si	Si	No
Año	1996	1999	1998	2003	1986

CRISP-DM (Cross Industry Standard Process for Data Mining) es otro proceso utilizado en la actualidad en proyectos de minería de datos [8, 13]. Este se caracteriza por dividir el proyecto en diferentes fases, tareas y actividades [14]. Otro proceso es Catalyst [8] conocido como P3TQ (Product, Place, Price, Time, Quantity) conformado por dos modelos [15, 16]: a) negocio (MII) y b) explotación de información (MIII). MII ofrece una guía para el desarrollo de un modelo de un problema u oportunidad de negocio y MIII proporciona una guía para la realización y ejecución de modelos de minería de datos [15, 16].

Por otra parte, un proceso industrial adaptado a la minera de datos es Six Sigma, definido como un método organizado y sistemático para la mejora de procesos, nuevos productos y servicios basados en métodos estadísticos y científicos con el fin de reducir las tasas de defectos establecidos por el cliente [17]. Six Sigma involucra el análisis de datos, a través de herramientas estadísticas, con el fin de reducir la variación mediante la mejora continua [18]. En la Tabla 1 se presenta un resumen de las principales características de los procesos presentados, los cuales en las últimas décadas han tenido un aumento importante, todos con el propósito de cumplir con los objetivos y requerimientos definidos en los proyectos.

A pesar de que estos procesos cumplen con el objetivo principal de guiar el descubrimiento de patrones de interés en volúmenes de datos, aún carecen de aspectos importantes como una mayor participación del usuario en cada una de las etapas y la presentación eficiente de los patrones de datos obtenidos. Ambos aspectos son fundamentales para una mejor explicación y entendimiento en la generación del nuevo conocimiento. En este mismo sentido, surge la necesidad natural de hacer una minería de datos centrada en el usuario con el propósito de mejorar la experiencia de los usuarios en el proceso de explotación de datos. En este sentido, un enfoque centrado en el usuario proporciona una mejora en la eficacia, satisfacción de usuario y accesibilidad.

3. Minería de datos centrada en el usuario

En este trabajo se proyecta una minería de datos centrada en el usuario con el propósito de mejorar la experiencia de usuario y tener proyectos funcionales y usables, teniendo como característica la creación de herramientas interactivas centradas en el usuario como apoyo para el proceso de la toma de decisiones. Para esto se incluyeron fundamentos del diseño centrado en el usuario a través de la norma ISO 9241-210:2010, y el proceso CRISP-DM. Se eligió CRISP-DM por ser uno de los principales procesos más utilizados por la comunidad internacional. Estudios recientes [13] destararon que CRISP-DM tiene una mayor aceptación con 43%, comparado con otros procesos, como SEMMA (8.5%) y KDD (7.5%). Se demostró también una alta aceptación de procesos propios, esto es, creados a la medida, con 27.5% de aceptación.

Por otro lado, se eligió la norma ISO 9241-210:2010, la cual es un estándar definido por la Organización Internacional de Normalización (ISO, por sus siglas en inglés), debido a las características, requisitos y recomendaciones que proporciona para el diseño centrado en usuario. Precisamente, estas características sirvieron como referencia para garantizar el diseño centrado en el usuario, a través de cinco etapas

iterativas [17]: a) análisis del contexto de uso, b) especificación de requerimientos, c) producción de soluciones de diseño, d) evaluación del diseño, y e) solución de diseño. En la Fig. 1 se presenta la estructura general de la minería de datos centrada en el usuario con base a los principios de la norma ISO 9241-210:2010 y CRISP-DM.

El objetivo del proceso es involucrar al usuario en etapas significativas de la minería de datos, siguiendo para esto un ciclo iterativo para conocer objetivos, necesidades, actividades, entornos de trabajo, entre otros aspectos, dividido en tres etapas principales (análisis, minería de datos y despliegue) y sus respectivas subetapas.

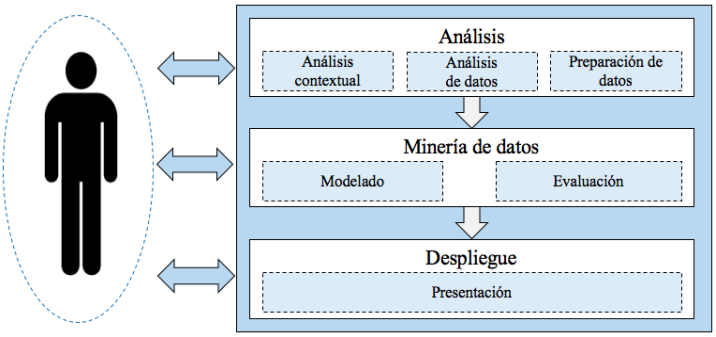

Fig. 1. Estructura general del proceso de minería de datos centrado en el usuario.

En la Fig. 2 se presentan las etapas y actividades que comprende el proceso de minería de datos centrado en el usuario, resaltando el diseño centrado en el usuario.

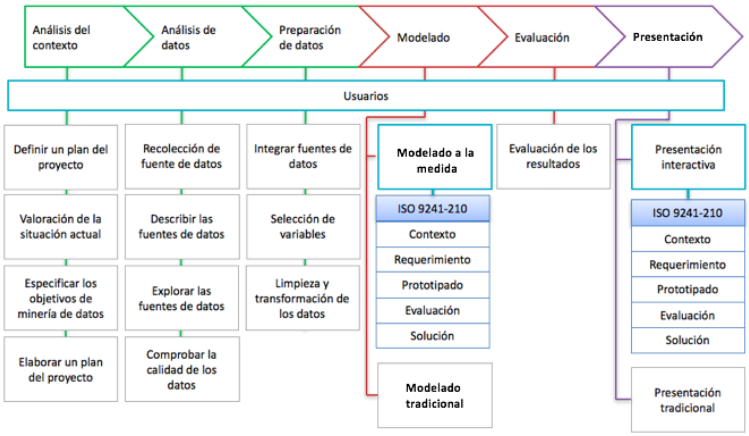

Fig. 2. Tareas generales comprendidas en el proceso de minería de datos centrado en el usuario.

De manera general, las fases que comprenden la minería de datos centrada en el usuario son: a) *análisis contextual*, que engloba en el entendimiento y descripción de los stakeholders, además, se definen los objetivos del proyecto y de minería de datos que se pretenden alcanzar y se elabora un plan general de proyecto; b) *análisis de datos*,

consiste en tener una aproximación sobre el entendimiento de los datos; c) *preparación de datos*, empleada para obtener la vista de datos minable sobre la cual se aplican las técnicas de minería de datos; d) *modelado*, contempla la selección de una o más técnicas de minería de datos para encontrar patrones de datos útiles, tendencias o nuevo conocimiento, en función de las necesidades del proyecto y del usuario (uso de herramientas libres o comerciales, o nuevas aplicaciones personalizadas); e) *evaluación*, consiste en evaluar los resultados desde el punto de vista de los objetivos del proyecto y de los usuarios; y f) *presentación*, comprende la presentación de los resultados obtenidos a través de interfaces interactivas.

4. Cáncer de mama en mujeres de origen mexicano

El cáncer de mama es un tumor maligno que se origina en las células de la mama. Estas células crecen de manera desordenada, logrando invadir tejidos que lo rodean, así como órganos distantes [18], representando en la actualidad una de las tres causas principales de muerte femenina en América Latina [19, 20]. La posibilidad de curación y de mejora en la calidad de vida de las pacientes con cáncer de mama depende de la extensión de la enfermedad en el momento del diagnóstico y de la aplicación adecuada de todos los conocimientos y recursos validados, incrementando la eficiencia y calidad técnica, utilizando para esto la evidencia científica [21]. En este sentido, surge la necesidad natural de propiciar investigaciones desde el punto de vista tecnológico y científico para desarrollar nuevas herramientas de apoyo que sirvan para identificar comportamientos y tendencias de la enfermedad.

Para esta investigación se utilizaron casos de cáncer de mama diagnosticados en mujeres de origen mexicano. La fuente de datos proviene del Programa de Vigilancia, Epidemiología y Resultados Finales (SEER, por sus siglas en inglés) del Instituto Nacional del Cáncer (NCI, por sus siglas en inglés). SEER se encarga de la recopilación de la información sobre los casos de cáncer diagnosticados, sobre las muertes atribuidas a esta enfermedad y la supervivencia de pacientes con cáncer.

En la actualidad, son diversas las investigaciones que se realizan a través del uso de los registros del cáncer, los cuales están a disposición de investigadores, médicos, funcionarios de salud pública, legisladores, políticos, grupos de investigación y público en general [20].

El análisis preliminar de datos identificó un total de 740506 registros y 146 variables, comprendidos entre 1973 y 2012, además, para esta investigación se utilizaron como referencia otros análisis realizados a la base de datos SEER. Estos análisis fueron efectuados bajo modelos matemáticos-estadísticos (análisis correlacional de datos y análisis de componentes principales) y la opinión de especialistas en el campo de la Salud, en los cuales se identificaron 34 variables significativas que tienen relación directa con el cáncer de mama y con registros suficientes en periodos consecutivos [22, 23].

A partir de este análisis se hizo una selección vertical (variables) y horizontal (registros) de los datos, se tomaron en cuenta únicamente variables asociadas al cáncer de mama en mujeres de origen mexicano. Así, la vista de datos minable final quedó

conformada por 16 variables y 2652 registros (Tabla 2), tomando como variable clase el estado de vida del paciente (Vital Status recode) cuyos valores binarios son: 0 para la mortalidad y 1 para la supervivencia.

Tabla 2. Variables seleccionadas para la vista de datos minable.

#	Nombre de variable	Descripción	Tipo
1	Marital Status at DX	Estado civil	Discreto
2	Age at diagnosis	Edad del paciente	Discreto
3	Month of diagnosis	Mes de diagnostico	Discreto
4	Year of diagnosis	Año de diagnostico	Discreto
5	Laterality	Lado donde se originó el tumor	Discreto
6	Behavior Code ICD-O-3	Tipo de comportamiento de la neoplasia	Discreto
7	Grade	Clasificación de las células cancerígenas	Discreto
8	Diagnostic Confirmation	Método de confirmación del cáncer	Discreto
9	Regional Nodes Examined	Numero ganglios linfáticos removidos y examinados	Discreto
10	RX Summ-Radiation	Método de radioterapia llevado a cabo	Discreto
11	RX Summ-Surg / Rad Seq	Secuencia de la cirugía y radioterapia	Discreto
12	Age Recode <1 Year olds	Grupo de edad(intervalos de 5 años)	Discreto
13	Survival Months	Tiempo de supervivencia del paciente(meses)	Discreto
14	Tumor Size	Tamaño del tumor	Discreto
15	AJCC Stage	Etapa de la enfermedad	Discreto
16	Vital Status recode	Estado de vida del paciente	Binario

Para este trabajo se desarrolló una aplicación (Fig. 3) con base en el diseño centrado en el usuario, la cual quedó integrada por cuatro secciones principales: a) Operadores, contiene funciones para cargar la fuente de datos, seleccionar la vista de datos minable, seleccionar el algoritmo de minería de datos y validar su precisión; b) Diseño de minería de datos, permite esquematizar una secuencia de operadores para la ejecución de los algoritmos de minería de datos c) Configuración de operadores, permite configurar los operadores en la sección de Diseño; y d) Resultados, presenta los resultados obtenidos a través de una interfaz interactiva.

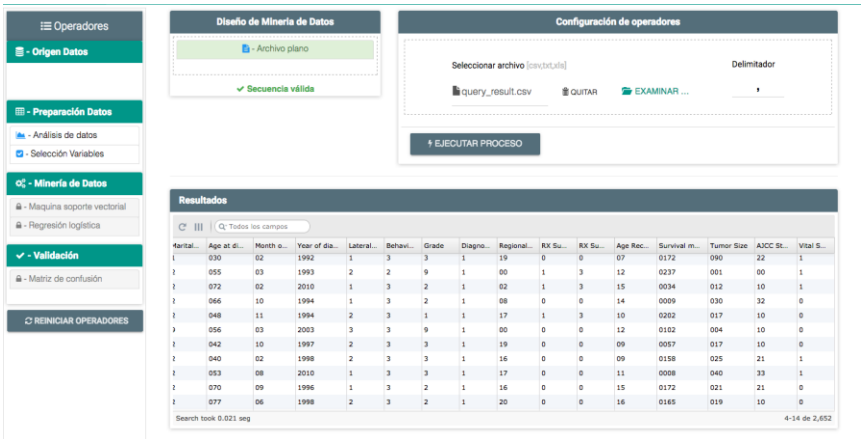

Fig. 3. Interfaz de la aplicación de minería de datos centrada en el usuario.

Por otra parte, con la finalidad de que el usuario cometa la menor cantidad de errores en el uso de la aplicación, se diseñó e implementó un autómata de usabilidad para

validar la secuencia correcta en la colocación de los operadores en la sección de Diseño. Para esto se definieron funciones para habilitar y deshabilitar los operadores con el propósito de mejorar la usabilidad y la experiencia del usuario. Los estados del autómata representan nodos secuenciales (Fig. 4).

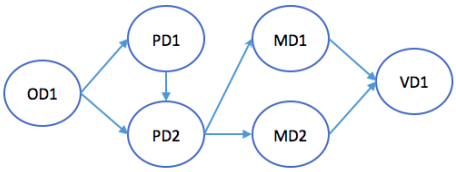

Fig. 4. Autómata para validar la ejecución de la secuencia de los operadores.

- Archivos planos (OD1). Tiene la funcionalidad de cargar y mostrar el conjunto de datos de tipo plano, con formatos csv, .txt o .xls.
- Análisis de datos (PD1). Permite analizar el conjunto de datos a través de tipos de gráficas diversas, por ejemplo, líneas, puntos, barras y otros.
- Selección de variables (PD2). Permite hacer una configuración de la entrada y salida de las variables que forman parte la vista de datos minable.
- Máquina de soporte vectorial (MD1). Con este operador se hacen las predicciones de una o más variables clase, tomando como entrada la vista de datos minable.
- Regresión logística (MD2). Este operador permite la predicción de una o más variables clase con base en las variables predictoras.
- Matriz de confusión (VD1). Mediante esta función se hace la evaluación de la precisión de los algoritmos de clasificación.

La validación de la secuencia de los operadores (Fig. 5 y 6) permitió guiar al usuario en el diseño de obtención de patrones, previniendo acciones fallidas. Por ejemplo, para el caso de la secuencia no válida, ésta se produce debido a que una vez cargado el conjunto de datos (OD1) es necesario definir las variables independientes, así como la variable clase (PD2), esto siguiendo el tipo de aprendizaje supervisado, y no utilizar antes el algoritmo de minería de datos (MD1 o MD2).

Cabe mencionar que una vez ejecutado algún algoritmo de minería de datos, es posible hacer una reconfiguración de las variables y conjunto de datos para obtener nuevos resultados. Es importante destacar también que a medida que se necesite incluir nuevos operadores a la herramienta, el autómata permite anexar nuevos nodos haciendo que el software sea escalable.

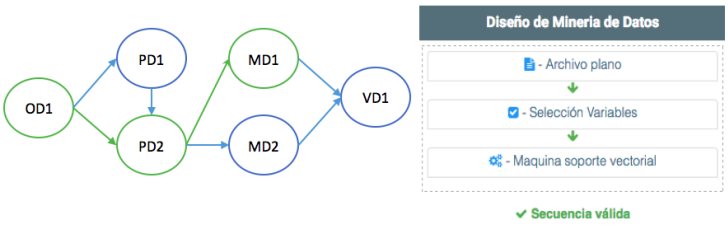

Fig. 5. Secuencia válida detectada por el autómata.

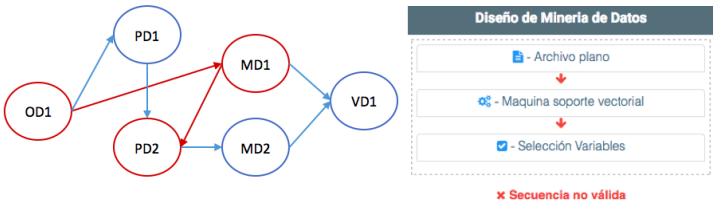

Fig. 6. Secuencia no válida detectada por el autómata.

Por otra parte, para ofrecer un mejor uso de la aplicación se realizaron pruebas de usabilidad con el propósito de hacer mejoras en la interfaz de usuario. Se trabajó con ocho usuarios con conocimientos en minería de datos. Se utilizó como método las tareas guiadas, esto es, se dictó a los usuarios en voz alta las tareas que debían realizar.

Los resultados alcanzados se muestran en la Fig. 7. Se identificó que seis usuarios concluyeron correctamente todas las tareas; otros dos tuvieron un error al completar una de las tareas asignadas. Se detectó además que en la etapa de selección de variables no había suficiente información para realizar la tarea. Esto permitió hacer las correcciones en la aplicación. Una vez construida la aplicación y evaluada desde el punto de vista de la usabilidad se hizo la ejecución para analizar los casos diagnosticados de cáncer de mama en pacientes mujeres de origen mexicano.

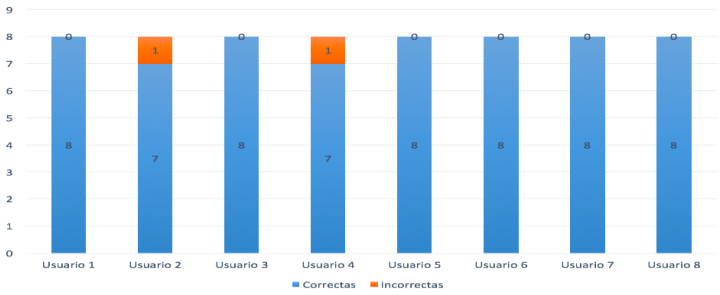

Fig. 7. Tareas correctas e incorrectas realizadas por los usuarios.

Posterior a la ejecución de los algoritmos, máquina de soporte vectorial y regresión logística, se evaluaron los resultados a través de una matriz de confusión mediante la cual se obtuvieron precisiones de clasificación de 87.4 y 85.6 %, respectivamente (Tabla 3). Se observó además que los resultados obtenidos para la supervivencia y mortalidad (Fig. 8 y 9) siguen un patrón similar al de los datos originales.

Tabla 3. Resultados de la precisión obtenidas por los algoritmos de clasificación.

Algoritmo	Total de casos	Correctos Positivos	Correctos Negativos	Falsos Positivos	Falsos Negativos	Precisión
Regresión logística	2,652	1695	624	239	94	87.4 %
Máquina de soporte vectorial	2,652	1725	547	316	64	85.6 %

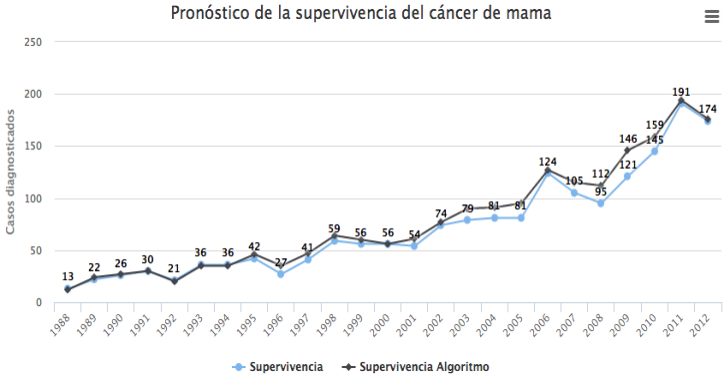

Fig. 8. Clasificación de la supervivencia del cáncer de mama.

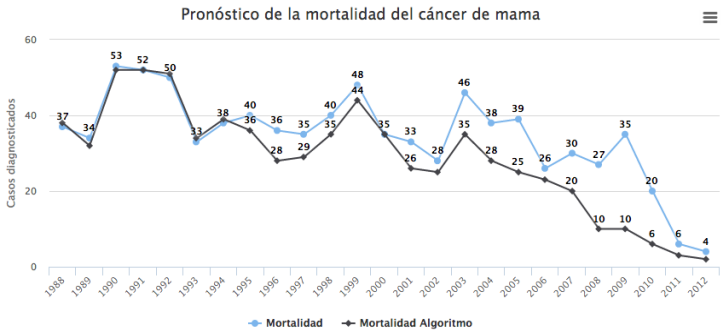

Fig. 9. Clasificación de la mortalidad del cáncer de mama.

Se calcularon otros parámetros de validez, como: sensibilidad y especificidad. La sensibilidad es la probabilidad de clasificar correctamente a un individuo vivo, es decir, la probabilidad de que una persona viva sea clasificada como un verdadero positivo (supervivencia), en este caso se obtuvo un valor de 95%, y la especificidad que es la probabilidad de clasificar correctamente a una persona muerta, es decir, una persona muerta sea clasificada como verdadero negativo (mortalidad), con 72%.

Para la presentación de los patrones de datos obtenidos se implementó una interfaz interactiva (Fig. 10), facilitando al usuario un mejor entendimiento de los resultados obtenidos a través de una serie de gráficas e interacciones definidas con los usuarios finales (médicos). Estas gráficas interactivas están relacionadas con las variables oncológicas de interés para los médicos especialistas del Hospital General La Raza (Cd. México). Estas variables son (descritas en la Tabla 2): Laterality, Behavior Code ICD-O-3, Grade, Diagnostic Confirmation y RX Summ-Radiation.

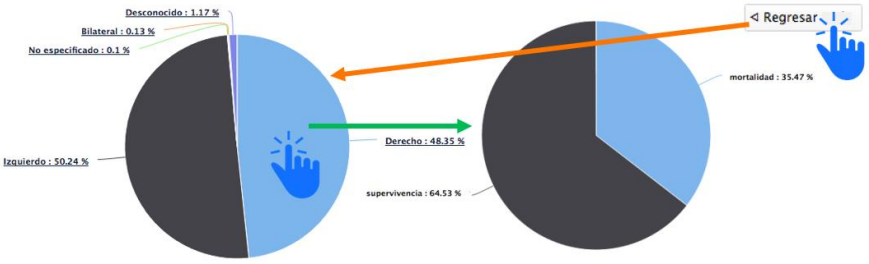

Fig. 10. Muestra de una gráfica interactiva de la lateralidad del tumor.

Una parte importante en el proyecto fue la evaluación de la presentación de los resultados a los médicos especialistas. Para esta evaluación fueron tomados en cuenta cuatro especialistas, quienes proporcionaron los requisitos para el desarrollo de la aplicación. La primera evaluación fue a través del método SIRIUS [24] obteniendo un porcentaje de usabilidad de 93.88%, 87.86%, 96.22%, 90.94%, respectivamente; y como segunda parte de evaluación se les presentó un checklist de verificación de usabilidad, sugerida en la Guía para Desarrollo de Sitios Web [25] con base a la heurísticas de Nielsen [26], obteniendo resultados con respecto a las siguientes heurísticas: a) navegación, todos los usuarios estuvieron de acuerdo con sus respuestas, logrando así una satisfacción positiva perfecta; b) visibilidad del estado del sistema, estética y diseño, retroalimentación, obtuvieron una satisfacción positiva alta de 3.25, 3.75 y 3 respectivamente; y c) ayuda ante errores, obtuvo una satisfacción baja con un valor de 1.25.

5. Conclusiones

La integración de un método centrado en el usuario en un proceso clásico de minería de datos no sólo fue para involucrar al usuario, sino también aspectos del diseño centrado en el usuario para el desarrollo de aplicaciones personalizadas, esto es, soluciones implementadas a la medida (Ad hoc). Esto se logró verificar mediante la experimentación práctica sobre la clasificación de la supervivencia y mortalidad de mujeres de origen mexicano diagnosticadas con cáncer de mama.

Como fase inicial del estudio se hizo la adquisición de la fuente de datos, correspondientes a registros clínicos de casos de cáncer de mama en mujeres de origen mexicano. Esta fuente de datos fue adquirida a través de un acuerdo de confidencialidad. Así, dada las características propias de la fuente de datos fue necesario hacer un tratamiento de ésta con el propósito de hacer una selección de datos significativos, vertical y horizontal. Producto de esto se generó una vista de datos conformada por 16 variables y 2652 registros.

Producto de la ejecución de los algoritmos implementados se obtuvo precisiones de 85.6% (máquina de soporte vectorial) y 87.4% (regresión logística). Esto indica que la clasificación de los casos de supervivencia y mortalidad asociados al cáncer de mama pueden ser pronosticados con una precisión notable, con una sensibilidad del 95% y

especificidad 72%, evidenciando la utilidad de los resultados obtenidos para el proceso de la toma de decisiones en el contexto médico.

Las pruebas de usabilidad realizadas sobre el prototipo, basadas en los métodos SIRIUS y Checklist, permitieron identificar mejoras sobre la interacción entre la aplicación y los usuarios.

Agradecimiento. Este trabajo forma parte del proyecto "Infraestructura para agilizar el desarrollo de sistemas centrados en el usuario" financiado por el Consejo Nacional de Ciencia y Tecnología, en el marco de Cátedras CONACYT (Ref. 3053).

Referencias

1. Zhao, Y., Chen, Y., Yao, Y.: User-centered interactive data mining. In: Cognitive Informatics, 5th IEEE International Conference, 457–466 (2006)
2. Horberry, T., Burgess-Limerick, R., Steiner, L.: Human Centred Design for Mining Equipment and New Technology. In: Proceedings 19th Triennial Congress of the IEA, 9, 14 (2015)
3. Brachman, R., Anand, T.: The process of knowledge discovery in databases. In Advances in knowledge discovery and data mining, American Association for Artificial Intelligence (1996)
4. Haun, S., Nürnberger, A.: Supporting exploratory search by user-centered interactive data mining. In: SIGIR Workshop Information Retrieval for E-Discovery (SIRE) (2011)
5. Habib ur Rehman, M., Liew, C. S., Wah, T. Y.: UniMiner: Towards a unified framework for data mining, Information and Communication Technologies (WICT), 134–139 (2014)
6. ISO 9241-210:2010.: Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems.International Standardization Organization (ISO) (2010)
7. Nigro, H. O., Gonzáles, S.E., Xodo, D. H.: Data Mining with Ontologies: Implementations. Findings, and Frameworks (2007)
8. Moine, J., Gordillo, S., Haedo, A.: Análisis comparativo de metodologías para la gestión de proyectos de Minería de Datos. Workshop Bases de Datos y Minería de Datos, 931–938 (2011)
9. SAS Institute.: Data Mining and the Case for Sampling. Data Mining Using SAS Enterprise Miner (2015), http://sceweb.uhcl.edu/boetticher/ML_DataMining/SAS- SEMMA.pdf
10. Sumathi, S., Sivanandam, S.: Introduction to Data Mining and its Applications. Studies in Computational Intelligence, 29, editado por Springer-Verlag, Heidelberg, Alemania (2006)
11. Peralta, F.: Elementos para un mapa de actividades para proyectos de explotación de información. Facultad Regional Buenos Aires, Argentina 52 (2013)
12. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRISP-DM 1.0 Step-by-step Data Mining Guide,74 (2000)
13. KDnuggets: Data Mning, Analytics, Big Data, and Data Science. http://kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html (2014)
14. Rivo, E., de la Fuente, J., Rivo, Á., García-Fontán, E., Cañizares, M., Gil, P.: Cross-Industry Standard Process for data mining is applicable to the lung cancer surgery domain. Clinical and Translational Oncology, 14(1), 73–79 (2012)
15. Pyle D.: Business modeling and data mining. Morgan Kaufmann 720 (2003)

16. Britos P. V.: Procesos de explotación de información basados en sistemas inteligentes. Universidad Nacional de la Plata, Buenos Aires, Argentina 234 (2008)
17. Jang, G., Jeon, J.: A Six Sigma Methodology Using Data Mining: A Case Study on Six Sigma Project for Heat Efficiency Improvement of a Hot Stove System in a Korean Steel Manufacturing Company. Research Topics on Multiple Criteria Decision Making, 72-80 (2009)
18. IMSS: Cáncer de mama. http://imss.gob.mx/salud-en-linea/cancer-mama (2016)
19. INEGI: Estadísticas a propósito del día mundial de la lucha contra el cáncer de mama. http://inegi.org.mx/saladeprensa/aproposito/2015/mama0.pdf
20. NCI: Cáncer de mama: Información general sobre el cáncer de mama. http://cancer.gov/espanol/tipos/seno (2016)
21. Secretaria de Salud: Diagnóstico y Tratamiento del Cáncer de Mama en Segundo y Tercer nivel de Atención (2009)
22. Molero G., Céspedes Y., Meda M.: Caracterización y análisis de la base de datos de cáncer de mama SEER-DB. IX Congreso Internacional Informática en Salud (2013)
23. Molero G.: Clasificador bayesiano para el pronóstico de la supervivencia y mortalidad de casos de cáncer de mama en mujeres de origen hispano (Tesis doctoral). Universidad de Guadalajara, México 155 (2014)
24. Suárez, M.: SIRIUS: Sistema de Evaluación de la Usabilidad Web Orientado al Usuario y basado en la Determinación de Tareas Críticas (2011)
25. Guiadigital.: Checklist de Usabilidad.http://guiadigital.gob.cl/articulo/usabilidad-0, (2016)
26. Nielsen, J.: Usability engineering. Academic Press Limited, Massachuetts, Estados Unidos, 361 (1993)