La minería de datos en apoyo a la toma de decisiones clínicas

Data Mining to Support Clinical Decision Making

Romero Zaldivar, Yidier; Ramírez Pérez, José Felipe; Soto Pelegrín, Lissette

Yidier Romero Zaldivar
yromero@uci.cu
UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS, Cuba
José Felipe Ramírez Pérez
jramirez14@uabc.edu.mx
UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA, México
Lissette Soto Pelegrín
lissettesp@uci.cu
UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS, Cuba

Revista Cubana de Transformación Digital
Unión de Informáticos de Cuba, Cuba
ISSN-e: 2708-3411
Periodicidad: Trimestral
vol. 3, núm. 2, 2022
rctd@uic.cu

Recepción: 03 Septiembre 2021
Aprobación: 18 Agosto 2022

URL: http://portal.amelica.org/ameli/journal/389/3893437005/

Resumen: Las técnicas de minería de datos constituyen una herramienta que se debe tener presente cuando se realiza un análisis predictivo. En el área de la me- dicina clínica se aplican estas técnicas de minería de datos predictivas para apoyar la toma de decisiones de los médicos en el diagnóstico de enferme- dades, con vistas al pronóstico de supervivencia de los pacientes y también para sugerir tratamientos. Los autores de este trabajo se propusieron rea- lizar una revisión de la literatura en aras de identificar las tendencias en el tema, las técnicas más precisas en la tarea de predicción y su aplicación en la medicina clínica. Para ello se aplicó un método de revisión sistemática de la literatura (SLR, por sus siglas en inglés). Al culminar se identificaron tres criterios importantes para elegir un modelo efectivo en el análisis predictivo, utilizando datos clínicos: la representación del problema, el poder explicativo de su salida y la capacidad de adicionar conocimiento previo de los expertos del dominio.

Palabras clave: minería de datos, modelos de predicción, sistemas de in- formación clínica.

Abstract: *Data mining techniques are a tool to keep in mind when a predictive analysis is needed. In the area of clinical medicine, these predictive data mining techniques are applied to support decision making by physicians in the diagnosis of disea- ses, for the prognosis of patient survival and to suggest treatments. The authors of this paper set out to conduct a literature review to identify trends in the sub- ject, the most precise techniques in the task of prediction and its application in clinical medicine. A method of systematic literature review (SLR) was applied to comply with the proposed objective. At the end of the work, three important criteria were identified to choose an effective model for predictive analysis in cli- nical data: the representation of the problem, the explanatory power of its exit and the ability to add previous knowledge of the experts in the domain.*

Keywords: data mining, predictive models, clinical information systems..

INTRODUCCIÓN

Con frecuencia es un reto para los médicos elegir un tratamiento efectivo para los pacientes de cáncer. Por ello, existe una tendencia a la terapia personalizada y la inmunoterapia, ya que constituyen tratamientos con buena respuesta del paciente y baja toxicidad. En el caso de la inmunoterapia, la comunidad científica ha llegado a resultados muy favorables para los pacien- tes de cáncer, pero aún se declaran grupos de estos que

no responden al tratamiento como se espera. Persiste entonces la incógnita de cuáles son las características de los pacientes, que les permiten responder de manera favorable a la inmunoterapia.

El crecimiento del volumen de datos asociados a tratamientos de pacientes, pronósticos y

diagnósticos de enfermedades, han propiciado la aplicación de la minería de datos como alter- nativa al análisis de los casos y la toma de decisiones clínicas. El término "minería de datos" aparece cada vez más en la literatura médica. Resulta una disciplina que engloba la estadís- tica, la inteligencia artificial y las tecnologías de bases de datos. El propósito de la minería de datos es obtener patrones comprensibles a partir de gran cantidad de datos, a través de técni- cas de análisis de datos como el aprendizaje automático y la estadística. Hace dos décadas se experimentó un incremento de la aplicación de esta minería de datos en datos médicos. En la medicina clínica hay varias aplicaciones prácticas que se pueden beneficiar de algunas solu- ciones de la minería de datos, las cuales posibilitan el modelado predictivo, explorando el co- nocimiento disponible en el dominio clínico, y se explican las decisiones propuestas una vez que se hayan empleado los modelos para el apoyo a las decisiones clínicas.

El objetivo de la minería de datos predictiva en la medicina clínica, es derivar modelos que

puedan utilizar la información específica del paciente, para predecir una respuesta de interés y dar soporte a la toma de decisiones clínicas. Los métodos de minería de datos predictivos pueden aplicarse en la construcción de modelos de decisión para procedimientos, como el pronóstico, el diagnóstico y la planificación del tratamiento, los cuales podrán integrarse a los sistemas de información clínica una vez que hayan sido evaluados y comprobados.

El propósito de este artículo es realizar una revisión metodológica de la minería de datos enfocada al proceso de análisis de datos y resaltar algunos de los aspectos más importantes relacionados con su aplicación en la medicina clínica. El alcance se limita a los métodos de la minería de datos predictiva, porque son los más consolidados en la práctica y que mejor se adaptan a los problemas surgidos del análisis de datos clínicos y el apoyo a decisiones clínicas.

METODOLOGÍA

Para el desarrollo de la investigación se utilizó el método de revisión sistemática de la litera- tura (SLR, por sus siglas en inglés) propuesto por autores como Xiao & Watson (2017) y To- rres-Carrión et al. (2018). Las fuentes utilizadas para obtener información fueron las bases de datos IEEE Xplore, Science Direct, ACM Digital Library y SciELO, y los motores de indexa- ción Google Scholar y Scopus.

Se utilizó como rango de fechas desde 2008 hasta 2018. Se realizó una búsqueda escalo- nada sobre la base al nivel de refinamiento: primero se fue decantando a partir del título de los trabajos, luego se revisaron los resúmenes del conjunto de artículos resultantes del paso anterior y, por último, se procedió a leer el texto completo de los artículos considerados de interés para los autores. Los criterios de búsqueda adoptados fueron basados en la aplicación de las técnicas de minería de datos , en apoyo a la toma de decisiones médicas (especialmente en la parte clínica). En la decantación de artículos se tuvo en cuenta el contexto nacional y las posibilidades con que cuenta Cuba para aplicar estas técnicas en un entorno real.

MINERÍA DE DATOS PREDICTIVA

Los métodos de minería de datos predictivos surgieron de distintos campos de la investiga- ción y a menudo utilizan soluciones de modelado muy diversas. Los siguientes son algunos criterios de comparación para cada método:

- • El tratamiento de datos incompletos y del ruido.
  • El tratamiento con distintos tipos de atributos (categóricos, ordinales y continuos).
  • La presentación de modelos de clasificación que permitan (o no) a los expertos en el do- minio examinarlos y conocer su funcionamiento (o razonamiento) interno.

• La reducción de la cantidad de pruebas, que se traduce en la cantidad de atributos nece- sarios para arribar a una conclusión.

• El costo computacional por la inducción y el uso de los modelos de clasificación.

• La habilidad para explicar las decisiones alcanzadas cuando es utilizado en la toma de decisiones.

• La generalización, que significa la habilidad de desempeñarse efectivamente frente a ca- sos desconocidos.

A continuación se muestran los métodos de minería de datos predictivos más utilizados en la actualidad.

Árboles de decisión

Los árboles de decisión utilizan el particionamiento de datos recursivos, lo cual induce cla- sificadores transparentes con un rendimiento que se ve afectado por la segmentación de los datos: las hojas de los árboles de decisión pueden tener muy pocos casos para obtener predic- ciones fiables. La complejidad computacional de los algoritmos de inducción es baja, debido a su potente heurística. Los paquetes más actuales para la minería de datos incluyen variantes del algoritmo de inducción de árboles de decisión C4.5 (Witten et al., 2017) y sus sucesores C5.0 y CART (Singh & Gupta, 2014). En la figura 1 se muestra un ejemplo de modelo de árbol de decisión en el cual se observa la capacidad expresiva del modelo.

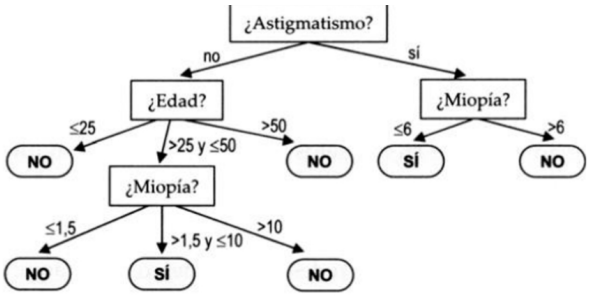

Figura 1. Ejemplo de modelo de árbol de decisión generado por el algoritmo C4.5.

FIGURA 1.
Ejemplo de modelo de árbol de decisión generado por el algoritmo C4.5.
sf

Regresión logística

La regresión logística es un método estadístico potente y bien establecido. Este es una extensión de la regresión tradicional, que puede modelar una salida binaria, la cual representa comúnmen- te la ocurrencia o no de algún evento (James et al., 2013). El modelo intrínseco para la probabili- dad es multiplicativo, similar al clasificador bayesiano simple, pero utiliza un método sofisticado basado en una estimación de la máxima verosimilitud para determinar los coeficientes en su fórmula de probabilidad (Yang, 2018). A este método se le dificulta el tratamiento de atributos con valores desconocidos. El modelo puede ser representado a través de un nomograma.

Redes neuronales artificiales

Las redes neuronales artificiales (RNA) es el algoritmo más popular de modelado de da- tos basado en inteligencia artificial, utilizado en la medicina clínica (Y. Zhang *et al*., 2016)

(Hsieh *et al*., 2018). Esto se debe a su buen desempeño predictivo, aunque pueden tener un conjunto de deficiencias como: la alta sensibilidad a los parámetros del método (incluidos aquellos que determinan la arquitectura de la red), el alto costo computacional en la fase de entrenamiento y la inducción del modelo que puede ser difícil de interpretar por los expertos del dominio (Mishra & Srivastava, 2014) (Biran & Cotton, 2017). Las redes neuronales pueden ser capaces de modelar las relaciones no lineales complejas, lo que constituye una ventaja so- bre métodos de modelado más sencillos, como el clasificador bayesiano simple y la regresión logística.

MáQuinas de vector soporte

Las máquinas de vector soporte (SVM, por sus siglas en inglés) es uno de los algoritmos de clasificación más potentes y precisos hoy en la predicción. Estos algoritmos están fundamen- tados en una base matemática robusta y en la teoría del aprendizaje estadístico. La esencia del método es un procedimiento que encuentra un hiperplano el cual separa los ejemplos en las distintas salidas (Abu Khousa & Campbell, 2012). Inicialmente, el método fue diseñado para resolver problemas de dos clases (como salida), donde el algoritmo encuentra un hiperplano con una distancia máxima al punto más cercano de las dos clases. Ese hiperplano se conoce como hiperplano óptimo y al conjunto de instancias que estén más cercanas al hiperplano óptimo se les denomina un vector soporte. La búsqueda del hiperplano óptimo proporciona un clasificador lineal (Wang et al., 2018).

Redes bayesianas

Las redes bayesianas son modelos gráficos probabilísticos, capaces de expresar conveniente- mente la unión de la distribución de probabilidades sobre un número de variables, a través de un conjunto condicional de distribuciones de probabilidades. Una red bayesiana es un grafo acíclico dirigido, donde cada nodo representa una variable estocástica y los arcos, la depen- dencia probabilística entre un nodo y su padre (no antecesor). Cada variable xi es asumida como independiente de sus no descendientes y tienen un conjunto de padres *pa(xi)*. Bajo esta

suposición, conocida como suposición de Markov (Ye, 2014), se une la distribución de proba- bilidad de todas las variables (*x*), que puede escribirse como la regla cadena siguiente:

$$p(x) = \prod_{i=1}^{n} p(x_i|pa(x_i)) \qquad (1)$$

sn
sf

La red es completamente especificada por un conjunto de distribuciones de probabilidad condicional, las cuales cuantifican las relaciones cualitativas entre las variables expresadas por el grafo. Como las distribuciones de probabilidad dependen de un conjunto de parámetros θ, como son las entradas de las tablas de probabilidad condicional por variables discretas o la media y la varianza de la distribución Gaussiana para variables continuas (Bielza & Larrañaga, 2014).

Las redes bayesianas pueden aplicarse en los problemas de clasificación, donde pueden ser vistas como una generalización del clasificador bayesiano simple por el modelado de las interacciones entre las variables del problema.

K-vecinos más cercanos

El algoritmo de k-vecinos más cercanos (kNN, por sus siglas en inglés) está inspirado en la solución que a menudo toman los expertos del área para tomar las decisiones basadas en ca- sos similares vistos con anterioridad. Para obtener una instancia de los datos, el clasificador kNN busca las k instancias (del grupo de entrenamiento) más similares y las clasifica en base a la clase predominante (Z. Zhang, 2016). La búsqueda por instancias más similares puede ser lenta y requiere de recorrer el conjunto de entrenamiento completo en el momento de la clasificación (Witten *et al*., 2017).

A modo de resumen parcial, los métodos predictivos mencionados anteriormente están

implementados en las herramientas de minería de datos actuales, ya sean solos o en combi- nación con preprocesamiento y funcionando con suficiente rapidez. La gran diferencia con el tratamiento de datos clínicos puede encontrarse en el desempeño predictivo e interpretativo de los resultados. A lo largo del trabajo se asume que los dos aspectos anteriores son impor- tantes y si algunos métodos poseen valores similares de

precisión en la predicción, se elige el modelo más fácil de interpretar y que ofrezca la mejor explicación de su respuesta.

La contribución de los datos al modelado predictivo en la medicina clínica

Los modelos predictivos en la medicina clínica son herramientas para ayudar la toma de de- cisiones que combinan dos o más características de los datos del paciente para predecir una respuesta clínica. Muchos modelos pueden utilizarse en los contextos clínicos por los médi- cos (Hurria et al., 2016) y pueden despertar una reacción de alerta frente a situaciones desfa- vorables (Rabbi et al., 2017). La minería de datos puede contribuir efectivamente al desarrollo de modelos predictivos clínicamente útiles, gracias a tres aspectos:

1. 1. Una solución comprensiva y enfocada a un propósito, para el análisis de datos que inclu- ye la aplicación de métodos y las soluciones derivadas de diferentes áreas científicas.
   2. La capacidad explicativa que poseen estos modelos.
   3. La capacidad de utilización del conocimiento previo del dominio en el proceso de análi- sis de los datos.

La explicación de los resultados

La minería de datos incluye soluciones que pueden tener un doble rol: ser utilizadas para de- rivar una regla de clasificación y para entender qué información está contenida en los datos disponibles. Inspirado en los primeros sistemas expertos, como MYCIN e INTERNIST (Duda & Shortliffe, 1983), los cuales eran utilizados en las aplicaciones médicas, la comunicación del

conocimiento explícito descubierto desde los datos y la explicación siguiente de las decisiones cuando el conocimiento es utilizado en la clasificación de un nuevo caso, es que se hace énfasis en el número de técnicas de minería de datos. En otros trabajos se muestra cómo los árboles de clasificación pueden revelar patrones interesantes en los datos (Phakhounthong *et al*., 2018).

Otro formalismo para representar los modelos de clasificación, que permiten una fácil ex-

plicación de los resultados, son las redes bayesianas. Afeni et al., (2017) publicó una aplicación interesante de las redes bayesianas, que utiliza el aprendizaje en el contexto de la minería de datos predictiva. En ese trabajo se compara la precisión entre el Naïve Bayes en el pronóstico de la hipertensión arterial, con tres redes bayesianas distintas introducidas desde los datos. La estructura permanece fija después de la fase de aprendizaje. El gráfico de salida solo muestra suposiciones a priori en las relaciones de variable, mientras que el conocimiento aprendido está oculto en las tablas de probabilidades.

La utilidad del conocimiento del dominio

Algunos algoritmos de minería de datos tienen una característica llamada "conocimiento pre- vio", que es tenida en cuenta junto a la capacidad de explicarse. El término "conocimiento pre- vio" se trata de información que es esencial para entender el problema. En el proceso de cons- trucción del modelo predictivo, utilizar el conocimiento previo significa estar capacitado para tener en cuenta la información conocida a priori y que no tiene que ser descubierta a partir de los datos. Este aspecto puede ser particularmente importante en el análisis de datos médicos.

El conocimiento previo puede ser expresado en diferentes formatos: existen ejemplos en las áreas de las reglas de decisión (Blobel, 2017), los modelos bayesianos (Krypotos et al., 2017), los conjuntos borrosos (Dutta, 2017) y la jerarquía de conceptos (Han et al., 2011). Un método que puede ser apropiado para tratar y codificar el conocimiento previo, son las redes bayesianas, donde el conocimiento previo es explotado para definir la estructura de red, díga-

se el número de variables, arcos y direcciones de los arcos. Además, siguiendo el paradigma bayesiano de las probabilidades a priori en las tablas de probabilidad condicional, son tenidas en cuenta en la relación entre las variables. Estas probabilidades permiten un modelo que sea derivado, siempre que la información proveniente de los datos sea débil y pueda ayudar a evi- tarse el sobreajuste, donde el modelo derivado puede reproducir los datos más cercanos y el error de clasificar correctamente los nuevos casos no conocidos.

El conocimiento previo puede ser aprovechado en la construcción de reglas de clasifica- ción: por ejemplo, un conjunto incompleto de reglas de clasificación proporcionado por los expertos puede refinar y argumentar en las bases de los datos disponibles, mientras la regla de búsqueda puede estar dirigida por cierto número de restricciones de monotonicidad.

DISCUSIÓN

Analizando la información obtenida de los trabajos revisados se identificaron varias áreas de aplicación de las técnicas de minería de datos predictivas. Entre ellas, el análisis de imágenes

médicas. Los autores Saranya & Satheeskumar (2016) utilizaron en su trabajo las técnicas de SVM, árbol de decisión (AD), Naïve Bayes, RNA y la regresión logística. Los mejores resulta- dos en cuanto a rendimiento y precisión los obtuvo el AD, seguido por la RNA y la SVM.

En otro grupo de trabajos se investigó el uso de las técnicas en la predicción de supervi- vencia en pacientes de cáncer. Los autores Pourhoseingholi *et al*., (2017) se enfocaron en pa- cientes con cáncer de colon y utilizaron las técnicas: SVM, AD (algoritmo C4.5), red de fun- ción de base radial (RBF), el K-NN, la red bayesiana y el modelo Naïve Bayes. En este trabajo se utilizaron métodos combinados o ensamblados, como *Bagging* y *Votting*.

Otros autores (Momenyan *et al*., 2018), enfocaron su trabajo en pacientes con cáncer de

mamas y utilizaron las técnicas AD (algoritmos CART, CHAID, QUEST, C5.0), las reglas de decisión (RD) deducidas de los árboles de decisión y la técnica de regresión logística. Esta úl- tima se decantó como la de mayor precisión en la mayoría de los análisis realizados.

Gupta *et al*. (2011) también trabajaron sobre el diagnóstico y la predicción del cáncer de

mamas y realizaron un estudio con las técnicas RNA, SVM, AD, NB. Estos autores comparan el modelo de RNA con el resto de los modelos y comprueban la potencialidad y efectividad de predicción del modelo frente al resto de los utilizados en el análisis.

Se identificaron otros trabajos, como el de Danjuma & Osofisan (2015), donde se aplican las técnicas de minería de datos para la diagnosis de Erythomatos-Escamosa. Para ello se uti- lizó el clasificador NB, la RNA (algoritmo perceptrón multicapa) y AD (algoritmo C4.5), don- de los mejores resultados se obtuvieron con las dos primeras técnicas.

Por último, mencionar un trabajo de revisión sistemática (Iavindrasana et al., 2009), el

cual detectó que las técnicas más publicadas en ese momento fueron: las redes bayesianas (RB), el clasificador Naïve Bayes (NB), las reglas de decisión (RD), el árbol de decisión (C4.5), la técnica de los k vecinos más cercanos (KNN), las redes neuronales artificiales (RNA) y las máquinas de vector soporte (SVM).

A modo de síntesis se elaboró la tabla 1, un resumen de las técnicas abordadas por los dis- tintos autores en sus trabajos.

Tabla 1. Técnicas aplicadas (o estudiadas) por autores ordenados ascendentemente por año.

	RL[1]	NB[2]	RB[3]	AD[4]	RD[5]	RNA[6]	KNN[7]	SVM[8]
Bellazzi & Zupan (2008)	x	x	x	x	x	x	x	x
Iavindrasana et al. (2009)		x	x	x	x	x	x	x
Gupta *et al.* (2011)		x		x		x		x
Danjuma & Osofisan (2015)		x		x		x		
Saranya & Satheeskumar (2016)	x	x		x		x		x
Pourhoseingholi *et al.* (2017)		x		x		x		x
Momenyan et al. (2018)	x			x				

Técnicas: [1] Regresión Logística, [2] Naïves Bayes, [3] Redes Bayesianas, [4] Árboles de Decisión, [5] Reglas de Decisión, [6] Redes Neuronales Artificiales, [7] K-Vecinos más cercanos, [8] Máquinas de Vector Soporte.

TABLA 1.
Técnicas aplicadas (o estudiadas) por autores ordenados ascendentemente por año.
Técnicas: 1 Regresión Logística, 2 Naïves Bayes, 3 Redes Bayesianas, 4 Árboles de Decisión, 5 Reglas de Decisión, 6 Redes Neuronales Artificiales, 7 K-Vecinos más cercanos, 8 Máquinas de Vector Soporte.

Se identificaron criterios importantes de selección de las técnicas de minería de datos predictivas. El método seleccionado debe ser capaz de inducir modelos de fácil comprensión,

puesto que los usuarios de la información generada no son precisamente expertos en técnicas de minería de datos. Esa información debe ser comprensible y los modelos explicar cómo se obtuvo la solución propuesta. Esto puede evidenciarse en la forma de representar el proble- ma y en el poder explicativo de la salida. A la hora de seleccionar un método se debe tener en cuenta la expresividad del modelo para representar el problema y describir la solución.

Otro criterio que se debe tener en cuenta es la participación del experto en el dominio, el médico que posee el conocimiento intrínseco y probado, el cual puede retroalimentar los métodos de minería de datos. En el área de la medicina, los expertos poseen un conocimiento previo que debe ser utilizado para reducir los espacios de búsqueda, y dar robustez y precisión a los métodos empleados.

Según los criterios anteriores se propone priorizar los métodos de árbol de decisiones, las reglas de decisiones, las redes bayesianas y las máquinas de vector soporte, por las caracterís- ticas que poseen para expresar el conocimiento y por su acertada aplicación en experiencias anteriores. Se debe tener en cuenta que ninguna técnica es idónea para resolver todo tipo de problemas, sino que se debe analizar el contexto del problema que se desea resolver, la natu- raleza de los datos y luego elegir la técnica que mejor se comporte frente a ese problema.

CONCLUSIONES

Con este trabajo, aplicando el método SLR, se identificó un conjunto de investigaciones que evidencian el interés de aplicar técnicas de minería de datos predictiva en el análisis de infor- mación clínica. Se expuso la esencia de los principales modelos referenciados en la literatura por su precisión de predicción. Por último, se identificaron tres criterios importantes para elegir un modelo efectivo en el análisis predictivo en datos clínicos: la representación del pro- blema, el poder explicativo de su salida y la capacidad de adicionar conocimiento previo de los expertos del dominio. Como trabajo futuro se prevé la elaboración de una metodología para apoyar la adopción de estas técnicas en el desarrollo de soluciones de soporte a la toma de de- cisiones clínicas.

REFERENCIAS

AbuKhousa, E., & Campbell, P. (2012). Predictive data mining to support clinical decisions: An overview of heart disease prediction systems. *2012 International Conference on In- novations in Information Technology (IIT),* 267-272. Obtenido de https://doi.org/10.1109/ INNOVATIONS.2012.6207745

Afeni, B. O., Aruleba, T. I., & Oloyede, I. A. (2017). Hypertension Prediction System Using Naive Bayes Classifier. *Journal of Advances in Mathematics and Computer Science,* 1-11. Obtenido de https://doi.org/10.9734/JAMCS/2017/35610

Bielza, C., & Larrañaga, P. (2014). Discrete Bayesian Network Classifiers: A Survey. *ACM Com- puting Surveys,* 47(1), 5:1-5:43. Obtenido de https://doi.org/10.1145/2576868

Biran, O., & Cotton, C. (2017). Explanation and justification in machine learning: A survey. *IJCAI-17 workshop on explainable AI (XAI)*, 8(1), 8-13.

Blobel, B. (2017). Knowledge representation and knowledge management as basis for decision support systems. *Int J Biomed Healthc*, 5, 13-20.

Danjuma, K., & Osofisan, A. O. (2015). Evaluation of predictive data mining algorithms in erythemato-squamous disease diagnosis. arXiv preprint arXiv:1501.00607.

Duda, R. O., & Shortliffe, E. H. (1983). Expert systems research. Science, 220(4594), 261-268.

Dutta, P. (2017). Decision Making in Medical Diagnosis via Distance Measures on Interval Valued Fuzzy Sets. *International Journal of System Dynamics Applications (IJSDA)*, 6(4), 63- 83. Obtenido de https://doi.org/10.4018/IJSDA.2017100104

Gupta, S., Kumar, D., & Sharma, A. (2011). Data mining classification techniques applied for breast cancer diagnosis and prognosis. *Indian Journal of Computer Science and Enginee- ring (IJCSE)*, 2(2), 188-195.

Hale, A. T., Stonko, D. P., Lim, J., Guillamondegui, O. D., Shannon, C. N., & Patel, M. B. (2018). Using an artificial neural network to predict traumatic brain injury. *Journal of Neurosur- gery: Pediatrics*, 23(2), 219-226. Obtenido de https://doi.org/10.3171/2018.8.PEDS18370

Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques. Techniques (3rd ed), *Morgan Kauffman*, 705. https://doi.org/10.1016/C2009-0-61819-5

Hsieh, M.-H., Hsieh, M.-J., Chen, C.-M., Hsieh, C.-C., Chao, C.-M., & Lai, C.-C. (2018). An Artificial Neural Network Model for Predicting Successful Extubation in Intensive Care Units. *Journal of Clinical Medicine*, 7(9), 240. Obtenido de https://doi.org/10.3390/jcm7090240

Hurria, A., Mohile, S., Gajra, A., Klepin, H., Muss, H., Chapman, A., Feng, T., Smith, D., Sun, C.-L., De Glas, N., Cohen, H. J., Katheria, V., Doan, C., Zavala, L., Levi, A., Akiba, C., & Tew, W. P. (2016). Validation of a Prediction Tool for Chemotherapy Toxicity in Older Adults With Cancer. *Journal of Clinical Oncology*, 34(20), 2366-2371. Obtenido de https://doi. org/10.1200/JCO.2015.65.4327

Iavindrasana, J., Cohen, G., Depeursinge, A., Müller, H., Meyer, R., & Geissbuhler, A. (2009). Clinical data mining: A review. Yearbook of medical informatics, 18(01), 121-133.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Lear- ning. Springer.

Krypotos, A.-M., Blanken, T. F., Arnaudova, I., Matzke, D., & Beckers, T. (2017). A Primer on Bayesian Analysis for Experimental Psychopathologists. Journal of Experimental Psycho- pathology, 8(2), 140-157. Obtenido de https://doi.org/10.5127/jep.057316