

Análisis de readmisión hospitalaria de pacientes diabéticos mediante aprendizaje computacional

Germán Cuaya-Simbro¹, Elias Ruiz¹, Angélica Muñoz-Meléndez²,
Eduardo F. Morales²

¹ Instituto Tecnológico Superior de Oriente del Estado de Hidalgo, Apan Hidalgo,
México

² Instituto Nacional de Astrofísica, Óptica y Electrónica, Tonantzintla, Puebla,
México

{gcuaya, eruiz}@itesa.edu.mx, {amunoz,emorales}@inaoep.mx

Resumen. La disminución de la tasa de readmisión hospitalaria temprana es una situación de interés debido a que esto generalmente es debido a complicaciones de alguna enfermedad en los pacientes. En México es particularmente importante la prevención de la readmisión temprana de pacientes diabéticos, puesto que esto está asociado a complicaciones derivadas de la enfermedad, como pueden ser problemas en los ojos, riñones, nervios, corazón, entre otros. Existen trabajos en donde se describen métodos computacionales para el diagnóstico de diabetes [10,4], pero en estos trabajos no se hace un análisis de los factores relacionados a las complicaciones de la enfermedad. En este artículo, se presenta un análisis basado en técnicas de aprendizaje computacional para estimar las variables más importantes asociadas a la readmisión hospitalaria en pacientes con diabetes. La metodología presentada contempla el uso de métodos de selección de atributos y selección de instancias mediante técnicas supervisadas y no supervisadas, así como la construcción de clasificadores con la información obtenida, finalmente, se presentan y se discuten los valores de precisión y recuerdo de los clasificadores para mostrar la efectividad de éstos en la predicción de readmisión temprana de pacientes diabéticos. Los resultados muestran las ventajas del uso de la metodología propuesta para la identificación de la información relevante asociada a la prevención de readmisión hospitalaria.

Palabras clave: selección de atributos, agrupamiento, clasificación, readmisión hospitalaria.

Hospital Readmission Analysis of Diabetic Patients through Computational Learning

Abstract. The decrease in the rate of early hospital readmission is a situation of interest because this is usually due to complications of some

disease in the patients. In Mexico, the prevention of early readmission of diabetic patients is particularly important, since this is associated with complications derived from the disease, such as problems in the eyes, kidneys, nerves, heart, among others. There are works describing computational methods for the diagnosis of diabetes [10,4], but in these works an analysis of the factors related to the complications of the disease is not done. In this article, we present an analysis based on computational learning techniques to estimate the most important variables associated with hospital readmission in patients with diabetes. The presented methodology contemplates the use of methods of selection of attributes and selection of instances by means of supervised and unsupervised techniques, as well as the construction of classifiers with the obtained information, finally, the values of precision and memory of the classifiers are presented and discussed. to show the effectiveness of these in the prediction of early readmission of diabetic patients. The results show the advantages of using the proposed methodology for the identification of relevant information associated with the prevention of hospital readmission.

Keywords: attributes selection, grouping, classification, hospital readmission.

1. Introducción

Actualmente los modelos de inteligencia computacional son usados cada vez con mayor frecuencia en el ámbito médico, como herramientas para el diagnóstico de enfermedades [6]. Sin embargo, es de especial interés usar las herramientas de inteligencia computacional en aplicaciones no solamente de diagnóstico, sino de predicción, en particular, es de suma importancia prevenir la aparición o complicación de enfermedades, lo cual generalmente implica una pronta (temprana) readmisión hospitalaria de un paciente.

En los hospitales es importante tener un índice bajo de readmisión hospitalaria [1] la cual es derivada de la complicación de algún padecimiento, como pueden ser cardiovascular, respiratorio, digestivo, diabetes, por mencionar algunos. Algunos países han creado programas para reducir la readmisión hospitalaria [12], con el objetivo de reducir los costos de salud y de mejorar la calidad de vida de los pacientes. Es importante señalar que, el reducir la readmisión hospitalaria implica la prevención y/o predicción de ésta.

La diabetes es una enfermedad con una incidencia cada vez más alta en el mundo [5,11]. En particular, en México es una enfermedad de alerta en hospitales [13]. Actualmente los hospitales se encuentran realizando esfuerzos en materia de diagnóstico preventivo hacia la población en general. Diagnosticar pre-diabetes o hacer recomendaciones de mejores hábitos de alimentación y ejercicio para pacientes con predisposición a ella tiene un impacto en la reducción de costos para las instituciones de salud y en una mejoría para la calidad de vida del paciente [2].

En contraste con lo anterior, poco se ha estudiado y discutido sobre aquellos pacientes diagnosticados con diabetes y los cuales tuvieron que ser reingresados al hospital en un corto tiempo por alguna complicación de la enfermedad, lo cual tiene implicaciones graves para la salud del paciente, en ocasiones la muerte. La mayoría de estas situaciones son ocasionadas debido a que un paciente llega a medicina general o a especialidades que no corresponden al tratamiento de la diabetes (endocrinología), lo que puede derivar en una complicación de la enfermedad como pueden ser, problemas en riones, pies, ojos o piel, corazón, entre otras. Dichas complicaciones pueden ser evitadas, pero se requiere un diagnóstico acertado y preventivo. Por este motivo, resulta importante contar con estrategias que permitan ayudar a predecir posibles complicaciones asociadas a la diabetes, lo cual generalmente deriva en una readmisión hospitalaria temprana, y en específico determinar los factores que están asociados a dicha readmisión.

En este trabajo de investigación se presenta un análisis de la información de una base de datos con registros de pacientes diagnosticados con diabetes los cuales fueron readmitidos en el hospital, determinando por medio de aprendizaje computacional aquellas variables que están asociadas a una readmisión temprana (en un periodo de tiempo menor a 30 días) y mostrando como el uso de este tipo de técnicas puede apoyar en la identificación de factores asociados a la complicación de enfermedades, en específico la diabetes.

2. Antecedentes

Debido a la importancia de la diabetes a nivel mundial, se han reportado diversos trabajos que utilizan técnicas de aprendizaje computacional para diagnosticar diabetes en pacientes [10,4]. La mayoría de dichos trabajos, utilizan reglas de decisión o clasificadores SVM (*Support Vector Machine*) en un enfoque supervisado. Sin embargo, su enfoque es principalmente el diagnóstico, no la predicción y/o prevención de las complicaciones de alguna enfermedad.

Por otro lado, existen trabajos que reportan el análisis de información para prevenir o predecir la readmisión de pacientes, como se presenta en [19] donde realizan un caso de estudio sobre una determinada raza para verificar la readmisión de pacientes pertenecientes a ésta en un lapso menor a 30 días, utilizando un enfoque basado en un algoritmo de redes bayesianas. En [20] hacen un análisis del impacto de la hemoglobina glicosilada con respecto de los niveles de reincidencia del paciente diabético al hospital a lo largo de varios meses, los autores concluyen que la probabilidad de readmisión está ligada al diagnóstico primario del paciente y a las acciones para mantener los niveles de hemoglobina glicosilada bajo control. Igualmente, en [14] muestran la importancia de tener un control glucémico al dar de alta al paciente para reducir el riesgo de reingreso en la población mexicana, aunque la mayor razón de reingreso fue debido a infecciones y enfermedades cardiovasculares. Los trabajos citados parten del análisis de información *a priori*, proporcionada generalmente por los expertos (en particular la consideraciones de variables a utilizar), en contraste al trabajo realizado en

esta investigación, se propone el análisis mediante aprendizaje computacional de los datos para determinar los factores o variables asociados a la complicación de una enfermedad la cual derive en la readmisión del paciente, en específico en pacientes diabéticos. La identificación correcta de los factores de riesgo asociados a complicaciones de cierta enfermedad propiciara un mejor pronóstico.

Existen diversas técnicas del aprendizaje computacional para analizar datos a fin de seleccionar atributos o variables de interés, seleccionar instancias o ejemplos, mejorar la clasificación a partir de datos entre otras, por ejemplo, [8,15]. Es por esto que en este trabajo se presentan resultados de la aplicación de diferentes algoritmos estándar de selección de variables, agrupamiento, y clasificación, para la identificación de factores de riesgo asociados a complicaciones de la diabetes, lo cual implica una readmisión temprana del paciente. Los métodos de selección de variable, de agrupamiento y de clasificación fueron aplicados a través Weka, el cual es un software para minería de datos que permite realizar clasificación y selección de atributos [7].

3. Metodología

En este trabajo se utilizó la base de datos descrita en [20]. Esta base de datos cuenta con 101,766 visitas de pacientes (registros) los cuales fueron hospitalizados y a quienes se les asignó un número de identificación. Cada paciente tiene registrado un número determinado de visitas, esto es, un paciente en esta base de datos puede tener registradas 4 visitas, y otro paciente puede tener asociado solo 1 visita. En una visita se toman la información de la Tabla 1 la cual cuenta con 55 atributos.

Tabla 1. Nombre y tipo de dato de los atributos de la base de datos. Información detallada de la tabla puede ser consultada en [20].

Atr.	Nombre del atributo	Tipo	Atr.	Nombre del atributo	Tipo
1	Encounter ID	Numérico	15	Num. of medications	Numérico
2	Patient number	Numérico	16	Num. of outpatient visits	Numérico
3	Race	Nominal	17	Num. of emergency visits	Numérico
4	Gender	Nominal	18	Num. of inpatient visits	Numérico
5	Age	Nominal	19	Diagnosis 1	Nominal
6	Weight	Numérico	20	Diagnosis 2	Nominal
7	Admission type	Nominal	21	Diagnosis 3	Nominal
8	Discharge disposition	Nominal	22	Number of diagnoses	Numérico
9	Admission source	Nominal	23	Glucose serum test result	Nominal
10	Time in hospital	Numérico	24	A1c test result	Nominal
11	Payer code	Nominal	25	Change of medications	Nominal
12	Medical specialty	Nominal	26	Diabetes medications	Nominal
13	Number of lab procedures	Numérico	27	24 features for medications	Nominal
14	Number of procedures	Numérico	28	Readmitted	Nominal

Cada visita registrada indica si el paciente fue readmitido antes de 30 días (clase 1), después de 30 días o ya no fueron readmitidos (clase 0), atributo *readmitted*. El análisis desarrollado identifica información relevante asociada a la readmisión temprana del paciente, y la relevancia de la información evaluada

mediante la construcción y evaluación de clasificadores, posteriormente se utilizan técnicas de agrupamiento para complementar o enriquecer la información relevante. Este proceso es descrito con detalle a continuación:

1. Se redujo el número de instancias mediante un enfoque supervisado de eliminación de registros incompletos.
2. Se realizó una conversión supervisada a datos numéricos de todos los atributos nominales.
3. Se ordenó la base de datos de acuerdo al atributo “Identificador de visita” (*Encounter ID*).
4. Se separó la base de datos de acuerdo al número de visitas de cada paciente, esto es, se creó una base de datos con la información de aquellos pacientes con únicamente 1 visita, M_1 , posteriormente se construye una nueva base de datos con la información de los pacientes que tuvieron al menos 2 visitas, M_2 , cabe señalar que la información contenida en esta nueva base de datos sólo está la información de la segunda visita de los pacientes dado que la información correspondiente a la primera visita ya se tiene almacenada en M_1 . Se continúa con este proceso hasta tener por separada la información de cada una de las visitas de los pacientes. Una vez acabado este proceso se obtuvo un total de 26 bases de datos, esto es, M_1, M_2, \dots, M_{26} , donde $|M_1| = 34321$, $|M_2| = 6484$, $|M_3| = 2259$, \dots , $|M_{26}| = 1$.
5. Se determinó trabajar solamente con la información de los pacientes que tienen al menos 3 visitas registradas, esto es, se usó sólo la información de M_1 , M_2 , y M_3 , debido a que el objetivo de este trabajo es prevenir las complicaciones de la diabetes, es decir, prevenir las readmisiones tempranas asociadas a esta enfermedad.
6. Se realizó un proceso de uniformidad de las bases de datos anteriores, buscando tener la misma información en cada una de ellas, tomando como referencias los elementos contenidos en M_3 , 2259, y eliminando los elementos diferentes en M_1 y M_2 . Así, $|M_1| = |M_2| = |M_3| = 2259$.
7. Se filtró la información de las bases de datos conservando únicamente aquellos pacientes (instancias) que fueron diagnosticados con algún tipo de diabetes, atributos *Diagnosis 1*, *2* o *3*. Así, $|M_1| = |M_2| = |M_3| = 485$.
8. Se unió la información de las 3 bases de datos anteriores en una sola de tal modo que se obtuvo la base de datos con la que se trabajó, BD_{all} , donde $DB_{all} = M_1 \cup M_2 \cup M_3$, $|DB_{all}| = 1455$.
9. Se realizó selección de atributos [8] mediante la regla $V_{sel} = \bigcup_{j=1}^n S_j$, donde S_j es el conjunto de variables obtenidas por un método de selección de atributos. En los experimentos $n = 5$ y el umbral de cada S_j fue fijado a los seis primeros resultados. Los métodos de selección de atributos utilizados fueron: método basado en correlación [9], método basado en correlación de *Pearson*, método basado en ganancia de información, y método basado en selección de atributos mediante el uso de reglas. Así el conjunto de variables V_{sel} obtenido tiene 13 elementos los cuales son mostrados en la Tabla 2. De acuerdo a las variables seleccionadas se creó una base de datos conservando únicamente dichas variables DB_{sv} .

Tabla 2. Variables escogidas por los métodos de selección de variables. Las variables con el número 27* corresponden a nombres de medicamentos consumidos por el paciente.

No. Atributo	Nombre del atributo	No. Atributo	Nombre del atributo
18	Number of inpatient visits	27*	glimepiride
22	Number of diagnoses	8	Discharge disposition
27*	Insulin	26	Diabetes medications
27*	glipizide-metformin	17	Number of emergency visits
27*	metformin-rosiglitazone	19	Diagnosis 1
10	Time in hospital	20	Diagnosis 2
27*	repaglinide		

10. Se procesó la información de DB_{all} con una técnica de aprendizaje no supervisado (k-medias de 2 grupos) para pesar de manera simple los ejemplos más fáciles de aprender y con ellos entrenar el modelo supervisado. Este procedimiento es una técnica simplificada de *boosting*. De lo anterior se obtienen 2 bases de datos, instancias correctamente aprendidas DB_{corr} y las incorrectamente aprendidas DB_{incorr} .
11. Finalmente se realizó clasificación de los datos utilizando técnicas de aprendizaje supervisado para cada conjunto de datos, tales como Redes Bayesianas [16], *bagging* [3], árboles [18] y vectores de soporte [17].

Un diagrama que ilustra el proceso de análisis de la información para determinar los factores relevantes asociados readmisión temprana de pacientes diabéticos a causa de complicaciones es presentado en las Figuras 1 y 2.

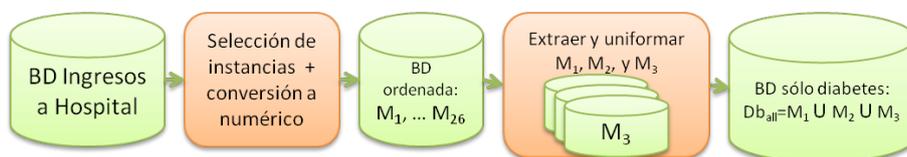


Fig. 1. Diagrama del pre-procesamiento de la información. Al final se consideran solamente pacientes con diagnóstico de diabetes.

4. Resultados y discusión

Para mostrar la efectividad del uso de la metodología propuesta en este trabajo se construyeron 4 clasificadores, utilizando las diferentes bases de datos generadas durante el proceso de análisis de la información, DB_{all} base de datos con todos los atributos, DB_{sv} base de datos con las variables de la Tabla 2, DB_{corr} base de datos con las instancias correctamente aprendidas en el proceso de agrupamiento.

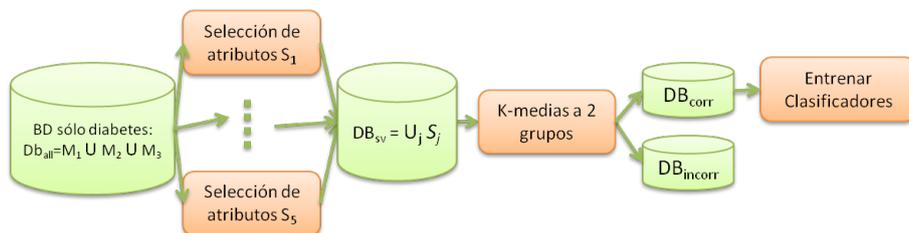


Fig. 2. Diagrama del proceso de análisis de la información el cual consiste en hacer selección de atributos, agrupamiento y clasificación.

Las métricas que se usaron para evaluar cada uno de los clasificadores fueron precisión, recuerdo y medida F.

La precisión se define como $\frac{VP}{VP+FP}$, donde VP son los verdaderos positivos (casos correctamente estimados) y FP son los falsos positivos (casos erróneamente clasificados como positivos). Esto es, es la proporción de instancias relevantes obtenidas respecto a las instancias recuperadas.

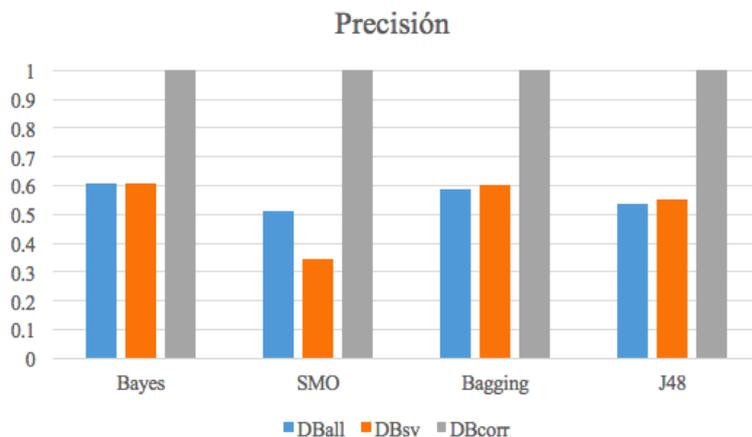


Fig. 3. Valor de precisión de los clasificadores con las diferentes bases de datos.

El recuerdo se define como $\frac{VP}{VP+FN}$, donde FN son los falsos negativos (casos erróneamente clasificados como negativos). Es decir, indica la proporción de instancias relevantes que se han recuperado sobre la cantidad total de instancias relevantes.

La medida F trata de establecer un compromiso entre la precisión y el recuerdo, evitando caer en una sobre representación de alguno de los dos. Se

define como se muestra en la Ecuación 1:

$$Medida F = \frac{2(precisión)(recuerdo)}{precisión + recuerdo} \quad (1)$$

Las Figuras 3, 4, 5 muestran los valores de las medidas de evaluación de los diferentes clasificadores construidos con las diferentes bases de datos.

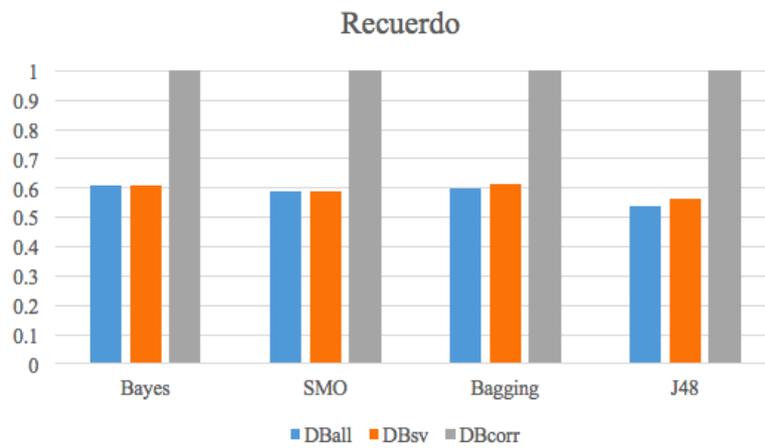


Fig. 4. Valor de recuerdo de los clasificadores con las diferentes bases de datos.

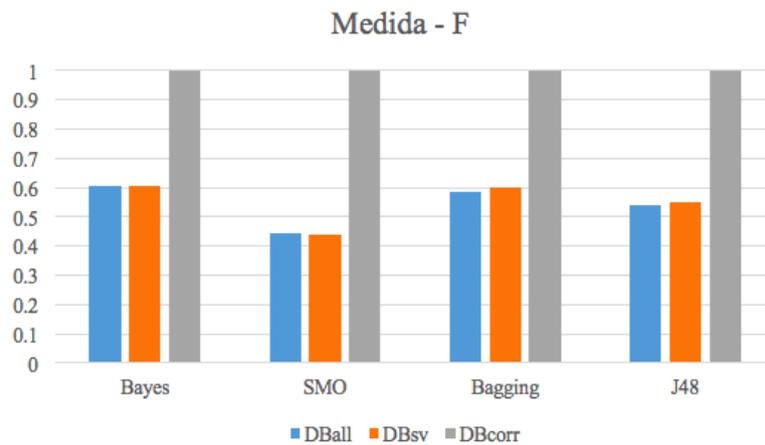


Fig. 5. Valor de la medida-F de los clasificadores con las diferentes bases de datos.

Como se puede observar los clasificadores construidos con las variables relevantes de la Tabla 2, tienen un desempeño similar a los clasificadores que fueron construidos utilizando todas las variables, esto es, se han determinado variables relevantes asociadas a la readmisión temprana debido a complicaciones de la diabetes.

Es de interés y de discusión los resultados obtenidos con la base de datos DB_{corr} , en principio esta base de datos es más pequeña que las bases DB_{all} y DB_{sv} , $|DB_{corr}| = 723$, el uso de agrupamiento permitió reducir el problema y analizar la información más relevante, debido a que las instancias consideradas en dicho conjunto de datos son muy similares y al momento de construir clasificadores con dicha información podemos determinar las relaciones y la relevancia de dichas variables, dado que obtenemos modelo más efectivos en términos de precisión, recuerdo y medida-F.

Con un análisis más detallado de los modelos creados con DB_{corr} permitió observar variables tales como la insulina la cual está estrechamente asociada a la detección de un riesgo de readmisión hospitalaria temprana, lo cual a partir de la simple selección de atributos no era posible determinar, así debe tenerse en cuenta que determinar subconjunto de forma supervisada ayuda en la explicación de las relaciones de las variables respecto a la clase de interés, readmisión.

Este análisis es complementario a la información mostrada en la Tabla 2 y debe de ser considerada por los expertos en el dominio para diseñar y conducir estudios más especializados en donde sólo se incorpore la información relevante y se descarte información que de acuerdo a los resultados no aporta información para la determinación de readmisión en un periodo menor a 30 días.

5. Conclusiones y trabajo futuro

Se presenta una metodología la cual permite la identificación de los factores relacionados a la readmisión hospitalaria temprana debido a complicaciones en pacientes diabéticos.

La selección de atributos en conjunto con el agrupamiento de instancias similares permite determinar las principales relaciones entre las variables y mejorar los modelos de prevención de readmisión hospitalaria temprana.

Se discuten las posibles ventajas del uso de la metodología propuesta desde el punto de vista computacional y además de las ventajas del uso de las variables relevantes desde el punto de vista clínico. Dando con ello un panorama de como técnicas de aprendizaje automático pueden apoyar en procesos de prevención médica, en este caso la prevención de readmisión temprana de pacientes diabéticos.

Como trabajo futuro, se extenderán el conjunto de pruebas para validar la efectividad de la metodología propuesta. Estas pruebas pueden involucrar el uso de todo el conjunto de datos, uso de datos filtrados por edad, entre otras. De la misma forma y con el objetivo de evaluar la efectividad de la metodología se probará ésta con diferentes bases de datos, tal como bases de datos médicas, climáticas o financieras, por mencionar algunas. Finalmente, con el objetivo de

medir la relevancia de la información se propone la validación de los resultados obtenidos, específicamente la selección de variables, con expertos en el dominio, para verificar si los resultados encontrados tienen validez desde el punto de vista clínico.

Referencias

1. Alonso Martínez, J.L., Llorente Díez, B., Echegaray Agara, M., Urbietta Echezarreta, M.A., González Arencibia, C.: Reingreso hospitalario en Medicina Interna. *Anales de Medicina Interna* 18, 28 – 34 (05 2001)
2. Barquera, S., Campos-Nonato, I., Aguilar-Salinas, C., López-Ridaura, R., Arredondo, A., Rivera-Dommarco, J.a.: Diabetes in Mexico: cost and management of diabetes and its complications and challenges for health policy. *Global Health* 9, 9 (2013), <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3599194/>
3. Breiman, L.: Bagging predictors. *Machine Learning* 24(2), 123–140 (1996), <https://doi.org/10.1007/BF00058655>
4. Fatima, M., Pasha, M.: Survey of machine learning algorithms for disease diagnostic. *Journal of Intelligent Learning Systems and Applications* 9, 1–16 (2017), www.scirp.org/journal/PaperInformation.aspx?PaperID=73781
5. Forouhi, N.G., Wareham, N.J.: Epidemiology of diabetes. *Medicine (Abingdon)* 42(12), 698–702 (Dec 2014), <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4282306/>, s1357-3039(14)00271-0[PII]
6. Foster, K.R., Koprowski, R., Skufca, J.D.: Machine learning, medical diagnosis, and biomedical engineering research - commentary. *Biomed Eng Online* 13, 94–94 (Jul 2014), <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4105825/>, 1475-925X-13-94[PII]
7. Frank, E., Hall, M.A., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I.H., Trigg, L.: Weka—a machine learning workbench for data mining. In: Maimon, O., Rokach, L. (eds.) *Data Mining and Knowledge Discovery Handbook*, 2nd ed., pp. 1269–1277. Springer (2010), https://doi.org/10.1007/978-0-387-09823-4_66
8. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. *Journal of Machine Learning Research* 3, 1157–1182 (2003), <http://www.jmlr.org/papers/v3/guyon03a.html>
9. Hall, M.A.: Correlation-based feature selection for discrete and numeric class machine learning. In: Langley, P. (ed.) *Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000)*, Stanford University, Stanford, CA, USA, June 29 - July 2, 2000. pp. 359–366. Morgan Kaufmann (2000)
10. Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., Chouvarda, I.: Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Journal* 15, 104 – 116 (2017), <http://www.sciencedirect.com/science/article/pii/S2001037016300733>
11. Mayer-Davis, E.J., Lawrence, J.M., Dabelea, D., Divers, J., Isom, S., Dolan, L., Imperatore, G., Linder, B., Marcovina, S., Pettitt, D.J., Pihoker, C., Saydah, S., Wagenknecht, L.: Incidence trends of type 1 and type 2 diabetes among youths, 2002–2012. *New England Journal of Medicine* 376(15), 1419–1429 (2017), <http://dx.doi.org/10.1056/NEJMoa1610187>, PMID: 28402773
12. McIlvennan, C.K., Eapen, Z.J., Allen, L.A.: Hospital readmissions reduction program. *Circulation* 131, 1796–1803 (May 2015), <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4439931/>

13. Meza, R., Barrientos-Gutierrez, T., Rojas-Martinez, R., Reynoso-Noveron, N., Palacio-Mejia, L.S., Lazcano-Ponce, E., Hernandez-Avila, M.: Burden of type 2 diabetes in Mexico: Past, current and future prevalence and incidence rates. *Preventive medicine* 81, 445–450 (2015), <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4679631/>
14. Molina-Corona, A., Zonana-Nacach, A.: Readmisin hospitalaria de pacientes diabeticos: seguimiento de 12 meses. *Revista Mdica del Instituto Mexicano del Seguro Social* 5, 469–474 (2010)
15. Olvera-López, J.A., Carrasco-Ochoa, J.A., Martínez Trinidad, J.F., Kittler, J.: A review of instance selection methods. *Artif. Intell. Rev.* 34(2), 133–143 (2010), <https://doi.org/10.1007/s10462-010-9165-y>
16. Pearl, J.: Probabilistic reasoning in intelligent systems - networks of plausible inference. Morgan Kaufmann series in representation and reasoning, Morgan Kaufmann (1989)
17. Platt, J.C.: Sequential minimal optimization: A fast algorithm for training support vector machines. Tech. rep., *Advances in Kernel Methods - support Vector Learning* (1998)
18. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann (1993)
19. Shameer, K., Johnson, K.W., Yahi, A., Miotto, R., Li, L., Ricks, D., Jebakaran, J., Kovatch, P., Sengupta, P.P., Gelijns, A., Moskovitz, A., Darrow, B., Reich, D.L., Kasarskis, A., Tatonetti, N.P., Pinney, S., Dudley, J.t.: Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: A case-study using Mount Sinai heart failure cohort. In: *Pacific Symposium on Biocomputing*. Pacific Symposium on Biocomputing. vol. 22, pp. 276–287 (2016), <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5362124/>
20. Strack, B., DeShazo, J.P., Gennings, C., Olmo, J.L., Ventura, S., Cios, K.J., Clore, J.N.: Impact of HbA1c measurement on hospital readmission rates: Analysis of 70,000 clinical database patient records. *BioMed Research International* 2014, 11 (2014), <http://dx.doi.org/10.1155/2014/781670>