

Aplicación de una red neuronal probabilística para predecir el desempeño académico de estudiantes de educación superior en línea

Rosa Leonor Ulloa Cazarez

Universidad de Guadalajara,
Sistema de Universidad Virtual, Guadalajara,
México

rosa.ulloa@udgvirtual.udg.mx

Resumen. La educación superior en línea (ESL) es una modalidad en crecimiento que presenta altos índices de reprobación. La predicción del desempeño del estudiante (DE) se ha utilizado para identificar oportunamente a los estudiantes en riesgo de reprobación, por lo que se propone la aplicación de una red neuronal probabilística (RNP) para clasificar el DE en cursos de ESL. Se utilizaron las calificaciones de los estudiantes de cuatro materias de ESL, que se prepararon para realizar dos experimentos, cada uno utilizando cuatro conjuntos de datos. La precisión de la RNP es comparada con la del Análisis Discriminante (AD). Los resultados mostraron mejor porcentaje de clasificación de la RNP que el AD en cinco de los ocho conjuntos de datos, mientras que en los otros tres resultaron iguales en sus porcentajes. Con base en ello, podemos concluir que una RNP es útil para clasificar el desempeño de un estudiante de ESL.

Palabras clave: Red neuronal probabilística, análisis discriminante, educación superior en línea, predicción del desempeño.

Probabilistic Neural Network to Predict the Academic Performance of Online Higher Education Students

Abstract. Enrolments of Online Higher Education (OHE) keep growing, but also its failure rates. The prediction of student academic performance (SP) is used to timely detect students at risk of failure. Thus, I propose to use a probabilistic neural network (PNN) to classify SP in terms of failure or success as a more understandable form to present prediction results. I use the scores of four subjects of an OHE program to execute two experiments, each one using the four sets of scores. I compared the accuracy of the PNN against the Discriminant Analysis (DA). The results show a better percentage of classification for the PNN than the DA in five of the eight datasets, while the rest of the three datasets obtained the same percentage for both models. On this basis, we can conclude that the PNN is useful to classify the academic performance of OHE students.

Keywords: Probabilistic neural network, discriminant analysis, online higher education, academic performance.

1. Introducción

La Educación Superior en línea es la modalidad instruccional con mayor crecimiento en el mundo [1] gracias a sus características de flexibilidad [2, 3] y a sus capacidades de adaptación y evolución. Alcanza estudiantes y profesores fuera de los contextos geográficos locales de las instituciones [4] y junto a este crecimiento de la matrícula, han surgido una serie de problemáticas relacionadas con el desempeño de los programas en línea en términos de la eficiencia terminal, la deserción y la reprobación de estudiantes [5-7].

En este contexto, la analítica educativa ha tomado prevalencia [8], en particular el uso de algoritmos de clasificación para predecir el desempeño del estudiante [9].

Algunos autores, con este propósito, exploran la aplicación de acciones gerenciales tales como la predicción temprana de estudiantes en riesgo de reprobación [10, 11] o de deserción [3]. También, se identifican trabajos evaluando el desarrollo de avanzadas herramientas tecnológicas para el monitoreo y acompañamiento de los estudiantes [12].

Se afirma que la aplicación de algoritmos de predicción en términos de clasificación guiados por una adecuada fundamentación teórica pedagógica y psicológica, proporcionan más elementos para la toma de decisiones por su facilidad de interpretación, que algoritmos con otro tipo de salidas, por ejemplo, una predicción numérica [13].

Sin embargo, se ha encontrado que la calidad de los datos utilizados es una de las deficiencias más preponderantes en la literatura sobre el tema [8, 14].

En trabajos previos, con el propósito de identificar tempranamente los estudiantes de ESL en riesgo [15, 16], se ha explorado el uso de algoritmos predictivos con salidas numéricas. Sin embargo, para continuar fortaleciendo la generación de conocimiento en el área de la analítica educativa, en este trabajo se propone un modelo predictivo en términos de clasificación, un método más comprensible para los profesionistas que no son del área computacional y que puede utilizarse de una forma más directa para sustentar la toma de decisiones y el despliegue de estrategias de retención y mejoramiento del desempeño de estudiantes [9].

En este trabajo utilicé datos de las calificaciones de los estudiantes en cuatro materias del primer semestre de un programa de Educación Superior en línea para predecir, clasificando, el desempeño de dichos estudiantes en cuanto a la aprobación o reprobación del curso.

El principal objetivo en este trabajo es explorar las ventajas de utilizar una Red Neuronal Probabilística (RNP) para predecir el desempeño de los estudiantes de Educación Superior en Línea, en términos de su precisión, facilidad de uso y de interpretación, y compararlo frente a un modelo estadístico común utilizado, el Análisis Discriminante (AD).

Se realizaron dos experimentos, uno utilizando el total de registros del conjunto de datos y otro con un subconjunto de estos datos. Para ello, se generó una RNP por cada conjunto de datos y se aplicaron en la predicción con un enfoque de clasificación, en lugar de predecir un valor numérico.

La precisión de la RNP (medida en porcentaje de estudiantes correctamente clasificados), es comparada con la del modelo estadístico, el AD.

Se formularon las siguientes hipótesis nula y alternativa a ser comprobadas:

- H_0 : El porcentaje del clasificador de una RNP es igual al del AD,
 H_1 : El porcentaje del clasificador de una RNP es diferente al del AD.

El RNP y el AD se generaron teniendo como variable predictora (independiente), a la calificación obtenida por los estudiantes en la primera unidad (U1) de cuatro cursos; la variable dependiente categórica, corresponde a una de las dos siguientes: “aprueba” y “no aprueba” en función de la calificación final de los mismos cursos, considerando que una calificación menor a 60 es igual a reprobar la materia.

El documento se desarrolla en las siguientes 4 secciones. La siguiente sección presenta la revisión de literatura sobre la predicción del desempeño del estudiante y sobre algoritmos de clasificación. La sección 3, describe la metodología para generar los modelos, la descripción de los datos utilizados, así como la teoría que sustenta a los dos modelos propuestos para la comparación, la RNP y el AD. La sección 4 detalla los resultados obtenidos y en la sección 5 se presenta la discusión y la identificación del trabajo futuro.

2. Trabajo relacionado

En la literatura se encuentra suficiente trabajo en cuanto a la predicción del desempeño del estudiante, implementando técnicas estadísticas y computacionales, con diferentes resultados y utilizando una variedad amplia de variables, que son relacionadas con una calificación final [9]. Esta variedad de resultados anima a continuar con esta línea de investigación, contribuyendo a la generación de conocimiento en las dos áreas, la de las ciencias computacionales para la predicción y la de la Educación Superior en línea.

Se identifican tres tipos de predicciones, la clasificación, la regresión y la estimación de la densidad [14]. Cada una aporta a la analítica educativa con una perspectiva rica para ser explorada, siendo la predicción con clasificadores, el método más comúnmente utilizado en educación [5]. También se encuentran trabajos que identifican estilos de aprendizaje y entrenan sistemas inteligentes de tutorías y para la evaluación del desempeño [13], así como la creación de modelos para predecir la deserción o retención de estudiantes [17].

En algunos artículos se analizan diversos factores relacionados con el contexto del estudiante y los vinculan con su desempeño académico [18]. Otros estudios analizan factores psicológicos, de la personalidad, y relacionados con prácticas de estudio y aprendizaje [19]. Y finalmente, hay estudios que utilizan los datos de las plataformas de gestión del aprendizaje [18, 20] como datos de entrada de los modelos.

Los algoritmos de clasificación más reportados en la literatura son C4.5, EM, Naïve Bayes, Support Vector Machines [18], algunas redes neuronales [21] y varios métodos basados en reglas y árboles de decisión [9]. No se encontró la aplicación de RNP para la predicción del DE lo cual justifica la elección de este método en el contexto específico de este trabajo.

Tabla 1. Descripción de los conjuntos de datos.

Materia	Conjunto	Matrícula total	Aprobados	Reprobados	Deserción
Pensamiento lógico-matemático (DPLM)	1	180	51	99	30
Fundamentos de Computación (CFC)	2	169	42	69	60
Planeación de proyectos (DPP)	3	165	45	56	64
Diagnóstico situacional (DS)	4	179	61	65	51
Totales		693	199	289	205

La predicción del desempeño del estudiante ha resultado en beneficios en cuanto a la prevención de la deserción y la emisión de señales de alerta de reprobación y en lo general, el uso de la analítica educativa ha probado ser de utilidad para fortalecer la toma de decisiones [8].

Sin embargo, resaltan dos características mencionadas en la literatura, mismas que se abordan en este trabajo como una contribución del mismo: la ya mencionada falta de claridad y calidad en el manejo, recolección y validación de los datos [8, 14, 18] y, la generación de un modelo predictivo que sea manejable tanto por los administradores educativos, como por los profesores y mentores [13, 19].

3. Metodología

3.1. Descripción de los datos

En referencia a uno de los propósitos en este trabajo sobre facilitar el uso y la aplicación de este modelo para los profesores y administradores educativos, se optó por trabajar con las calificaciones de los estudiantes que se pueden obtener mediante las herramientas de reporte que tienen prácticamente todas las plataformas de gestión del aprendizaje. En el caso que se presenta, se usaron los reportes de dos plataformas distintas, Moodle, y una plataforma de desarrollo propio de la institución de Educación Superior que oferta el programa de análisis, denominada Metacampus [22].

Se descargaron las calificaciones obtenidas en el total de materias cursadas (cuatro) por los estudiantes durante el primer semestre de un programa de ESL, mismas que conforman los cuatro conjuntos de datos descritos en la Tabla 1.

Trabajar con datos del primer ciclo escolar del estudiante, permite tomar decisiones en tiempo oportuno, así como el despliegue de estrategias remediales. En ambos casos de plataformas, las calificaciones finales son calculadas automáticamente, ya sea obteniendo el promedio general o la sumatoria de los valores ponderados de cada actividad en el curso. Los puntajes obtenidos por los estudiantes en las actividades, son otorgadas manualmente por los profesores de cada curso, considerando rúbricas con criterios de evaluación especificados de antemano por una academia [23]. El valor de la calificación de la unidad 1, es calculado manualmente por la autora de este trabajo, como el promedio (Pensamiento lógico matemático-DPLM, Planeación de proyectos-

DPP, y Diagnóstico situacional-DS) o la sumatoria (Fundamentos de Computación-CFC) de las calificaciones de las actividades que tuvieron lugar en el periodo de la unidad en congruencia con el sistema de evaluación establecido para cada curso.

Se realizaron dos experimentos, uno, utilizando el total de registros de la base de datos, siendo esta BD1, con 693 registros. En el segundo, se eliminaron los registros de los estudiantes que desertaron del curso, creando una segunda base de datos, denominada BD2, con un total de 488 registros.

3.2. Red neuronal probabilística

Una RNP es una red neuronal a la que se le reemplaza la función de activación sigmoidea y ejecuta un método no paramétrico para clasificar observaciones con bastante confiabilidad en la clasificación y una velocidad de respuesta que puede ser superior al mecanismo de propagación inversa [24]. En nuestro caso, las observaciones corresponden a los estudiantes en línea dentro de g grupos o categorías (para nuestro estudio, $g = 2$, siendo “aprueba/no aprueba”). La RNP funciona con base en v variables cuantitativas, que para nuestro estudio $v=1$ es la variable independiente cuyo valor corresponde a la calificación obtenida por cada estudiante al término de la primera unidad del curso [25].

La RNP construye una estimación no paramétrica de cada función de densidad por cada uno de las g categorías a una posición deseada, con base en observaciones cercanas a esa categoría. La estimación es basada en una ventana de Parzen que pondera observaciones a partir de cada categoría de acuerdo a su distancia a partir de la posición especificada.

Una RNP consiste de cuatro capas, que son ilustradas en la Figura 1:

- 1) Capa de entrada – tiene v neuronas, donde $v = 1$ (una variable de entrada, que es la calificación de la primera unidad obtenida por cada estudiante);
- 2) Capa de patrones – con n neuronas, donde $n = 180$ patrones (que corresponde al número de estudiantes del curso 1)
- 3) Capa sumatoria – tiene g neuronas, donde $g = 2$ (relacionados con las dos clasificaciones: “aprueba” y “no aprueba”)
- 4) Capa de salida – contiene una neurona binaria para cada categoría g , que se activa o desactiva si determinado estudiante es o no asignado a la categoría correspondiente

En la Fig. 1, se diagrama uno de los conjuntos de datos, a fin de ilustrar el flujo de datos a través de la RNP, en el sentido de que la variable predictora v de la capa de entrada, es estandarizada con base en la media de sus valores, divididos entre su desviación estándar. Estos valores sirven de entrada a las neuronas de la capa de patrones, que los turna a través de una función de activación usando los valores de entrada para estimar la función de densidad de probabilidad para cada una de las g categorías a una posición dada. La función de activación usada es la expresada en la Ecuación 1, si el estudiante i pertenece a la categoría j , es decir, la Ecuación 1 sería igual a cero si el estudiante i no pertenece a la categoría j ($g_{ij}=0$), donde:

— X corresponde al conjunto de valores de la variable de entrada, es decir, al conjunto de calificaciones de la primera unidad obtenida por cada estudiante.

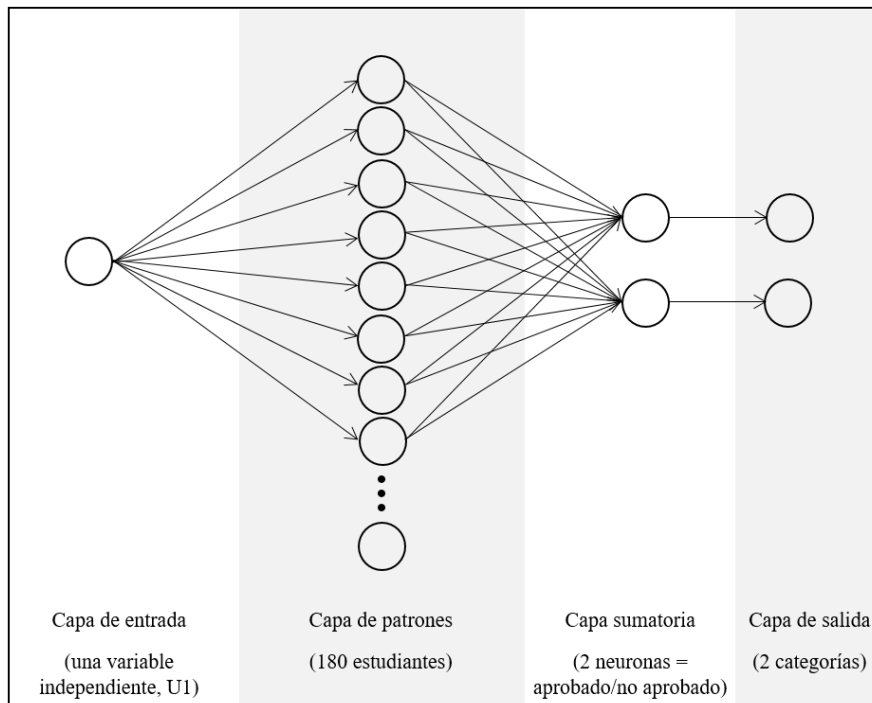


Fig. 1 Descripción del diagrama de una RNP, mostrando el primer conjunto de datos en nuestro estudio.

- x_i corresponde a la i -ésima calificación del i -ésimo estudiante.
- σ es la desviación estándar de las calificaciones del grupo:

$$g_{ij} = W\left(\frac{X - x_i}{\sigma}\right). \quad (1)$$

La función W corresponde a la Gaussiana expresada en la Ecuación 2, donde $\|X - x_i\|^2$ es la distancia euclidiana entre X y x_i :

$$\exp\left(-\frac{\|X - x_i\|^2}{\sigma^2}\right). \quad (2)$$

Estas estimaciones de densidad son turnadas a la capa sumatoria, misma que combina la información recibida con los datos de los n estudiantes con sus correspondientes costos tanto de probabilidad como de clasificación errónea para derivar un puntaje para cada una de las g categorías. Estos puntajes son turnados a las neuronas binarias de la capa de salida.

3.3. Análisis discriminante

En un análisis estadístico de regresión, la variable dependiente es numérica, mientras en el AD, la variable dependiente es categórica. Esta variable en nuestro estudio adquiere dos variantes o categorías: “aprueba” o “no aprueba”.

Las variables independientes son numéricas, correspondiendo a una sola en nuestro estudio, es decir, la calificación obtenida por cada estudiante al término de la primera unidad del curso, U1. Estas variables son útiles para determinar a qué categoría pertenecen los estudiantes.

El AD encuentra relaciones lineales entre las variables independientes que mejor discriminen a los estudiantes en las categorías. Enseguida, el AD construye una regla de decisión para asignar un estudiante nuevo a una de las categorías preestablecidas, es decir, una regla que responda a la pregunta “¿a qué categoría debería pertenecer cierto estudiante con determinada calificación obtenida en su primera unidad del curso?”

El objetivo en un modelo estadístico de AD es construir n combinaciones lineales de las v variables independientes tal que mejor discriminen entre las g categorías. La j -ésima función discriminante tiene la forma de la Ecuación 3, donde cada valor de Z es estandarizada, al igual que en la RNP, con base en la media de sus valores dividida entre su desviación estándar. Las funciones discriminantes se encuentran a partir de los eigenvalores de W^tB , donde W corresponde a la suma de cuadrados dentro de las categorías y la matriz de los productos cruz, mientras que B corresponde a la suma de cuadrados entre las categorías y la matriz de los productos cruz. Los coeficientes de las funciones discriminantes son obtenidos a partir de eigenvectores [26]:

$$D_j = d_{j1}Z_1 + d_{j2}Z_2 + \dots + d_{jv}Z_v. \quad (3)$$

Para clasificar un nuevo estudiante, las funciones de clasificación son derivadas y una puntuación se deriva para cada categoría. Una puntuación para la j -ésima categoría se calcula con base en la Ecuación 4; el nuevo estudiante es clasificado en una categoría teniendo el valor más grande para $C_j * \text{previo}_j$, donde previo_j corresponde a la probabilidad previa de pertenecer a la j -ésima categoría:

$$C_j = c_{j1}X_1 + c_{j2}X_2 + \dots + X_{jv}Z_v + C_{j0}. \quad (4)$$

4. Resultados

Tanto la RNP, como el AD fueron entrenados y probados mediante el método de validación *leave-one-out* (LOO), ello por tratarse de conjuntos de datos pequeños y porque el LOO es un método determinístico eliminando cualquier aleatoriedad en la selección de sus datos para ser entrenado [27].

En el caso del primer experimento, se utilizó la BD1 con un total de 693 registros. En el segundo experimento se utilizó la BD2 con un total de 488 registros. Los resultados por conjunto de datos para la RNP y el AD, se muestran en la Tabla 2.

Con base en los resultados, podemos rechazar la hipótesis nula en cinco de los ocho conjuntos de datos y aceptar la hipótesis alternativa en favor de la RNP formulada en la sección de Introducción de nuestro estudio:

Tabla 2. Resultados en la precisión de la RNP y AD por conjunto de datos, utilizando BD1 y BD2. El valor en columnas RNP y AD corresponde al porcentaje de estudiantes clasificados correctamente, el Tamaño corresponde al total de registros por materia utilizados).

Experimento	Materia	Tamaño	RNP	AD
1 – 693 registros totals	DPLM	180	81.67	72.22
	DPP	165	94.55	93.33
	DS	179	86.03	75.98
	CFC	169	85.80	85.80
2 – 488 registros totals	DPLM _b	150	80.67	73.33
	DPP _b	101	90.10	90.10
	DS _b	128	81.25	74.22
	CFC _b	109	79.92	79.92

H_1 : La precisión del clasificador de una RNP es diferente a la del AD.

Asimismo, se acepta la hipótesis nula para los conjuntos CFC, DPP_b y CFC_b:

H_0 : La precisión del clasificador de una RNP es igual a la del AD.

5. Conclusiones

Con base en los resultados de la Tabla 2 y las hipótesis aceptadas, podemos concluir que una RNP fue mejor o igual, (pero nunca peor) que un AD cuando se comparan en su precisión. Por tanto, una RNP puede ser aplicada para predecir con base en un enfoque de clasificación *aprueba/no aprueba*, el desempeño de un estudiante en programas de ESL.

Los dos modelos, RNP y AD generados, producen resultados predictivos que permiten la comprensión de la predicción y pueden ser utilizados para tomar acciones rápidas de intervención por un profesor o un mentor. Sin embargo, la rapidez de respuesta de la RNP así como su mayor precisión, la hacen una alternativa más viable que el AD.

Por otro lado, se usaron únicamente las calificaciones de los estudiantes para generar los modelos, lo que le da mayor simplicidad a la construcción del modelo siendo también más sencillo de replicar pues los datos son de relativo fácil alcance y estandarización.

Se resalta la generación de dos experimentos a partir de crear una segunda base de datos con las calificaciones de los estudiantes, eliminando a aquellos que desertaron (BD2). Sin embargo, se observó que la precisión de los dos modelos disminuyó cuando se usó BD2 respecto a los obtenidos con la BD1.

Estos resultados abonan al conocimiento en la generación y aplicación de algoritmos computacionales para predecir en términos de clasificación *aprueba/no aprueba*, y se habilita la replicación de este estudio bajo las condiciones estipuladas en la sección 3 de Metodología. Este tipo de predicciones en términos de clasificación, son útiles para

la toma rápida de decisiones por parte de docentes, mentores y administradores educativos, pues el algoritmo directamente interpreta la clasificación y la presenta de manera sencilla, tal que ayuda a identificar a los estudiantes en riesgo.

Para el trabajo futuro se incluye la aplicación de clasificadores basados en máquinas de soporte vectorial, cuyo desempeño será comparado con los obtenidos de la RNP y AD en este estudio.

Referencias

1. Clinefelter, D.L., Aslanian, C.B.: *Online college students 2016: Comprehensive Data on Demands and Preferences*. Learning House, Louiseville (2016)
2. García Aretio, L.: Historia de la educación a distancia. *Revista Iberoamericana de Educación a Distancia*, 2(1), pp. 8–27 (1999)
3. Lykourantzou, I., Nikolopoulos, V., Mpardis, G., Loumos, V.: Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education* (2009)
4. Kardan, A.A., Sadeghi, H., Ghidary, S.S., Fani-Sani, M.R.: Prediction of students course selection in online higher education institutes using neural network. *Computers & Education*, pp. 1–11 (2009)
5. Thakar, P., Mehta, A.: Manisha, Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue. *International Journal of Computer Applications*, 110(15), pp. 60–68 (2015)
6. Mohammadyari, S., Singh H.: Understanding the effect of e-learning on individual performance: The role of digital literacy. *Computers & Education*, 82, pp. 11–25 (2015)
7. Bawa, P.: Retention in online courses. Exploring issues and solutions. A literature review. *SAGE Open*, 6(1), pp. 1–11 (2016)
8. Banihashem, S.K., Aliabadi, K., Ardakani, S.P., Delaver, A., Ahmadabadi, M.N.: Learning Analytics: A Systematic Literature Review. *Interdisciplinary Journal of Virtual Learning in Medical Sciences*, 9(2) (2018)
9. Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., Murray, D.J., Long, Q.: Predicting academic performance by considering student heterogeneity. *Knowledge-Based Systems*, 161(1), pp. 134–146 (2018)
10. Castillo Merino, D., Serradell López, E.: An analysis of the determinants of students' performance in e-learning. *Computers in human behavior*, 30, pp. 476–484 (2014)
11. Romero, C., López, M.-I., Luna, J.-M., Ventura S.: Predicting students' final performance from participation in on-line discussion forums. *Computers & Education*, 68(1), pp. 458–472 (2013)
12. Chrysafiadi, K., Virvou, M.: PeRSIVA: An empirical evaluation method of a student model of an intelligent e-learning environment for computer programming. *Computers & Education*, pp. 322–333 (2013)
13. Gobert, J.D., Kim, Y.J., Sao Pedro, M.A., Kennedy, M., Betts, C.G.: Using educational data mining to assess students' skills at designing and conducting experiments within a complex systems microworld. *Thinking Skills and Creativity*, 18(1), pp. 81–90 (2015)
14. Avella, J.T., Kebritchi, M., Nunn, S.G.: Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review. *Online Learning*, 20(2), pp. 13–29 (2016)
15. Ulloa-Cazarez, R.L., Lopez-Martin, C.: Neural networks for predicting student performance in online education. *IEEE Latin America Transactions*, 16(7), pp. 2053–2060 (2018)

16. Ulloa-Cazarez, R., Lopez-Martin, L., Abran, C.A., Yañez-Marquez, C.: Prediction of Online Students Performance by Means of Genetic Programming. *Applied Artificial Intelligence*, 32(9-10), pp. 858–881 (2018)
17. Mason, C., Twomey, J., Wright, D., Whitman, L.: Predicting Engineering Student Attrition Risk Using a Probabilistic Neural Network and Comparing Results with a Backpropagation Neural Network and Logistic Regression. *Research in Higher Education*, 59(1), pp. 382–400 (2018)
18. Gašević, D., Dawson, S., Rogers, T., Gasevic, D.: Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *Internet and Higher Education*, 28, pp. 68–84 (2016)
19. Yeh, Y.-C., Kwok, O.-M., Chien, H.-Y., Sweany, N.W., Baek, E., McIntosh, W.: How College Students' Achievement Goal Orientations Predict Their Expected Online Learning Outcome: The Mediation Roles of Self-Regulated Learning Strategies and Supportive Online Learning Behaviors. *Online Learning*, 23(4) (2019)
20. Yang, T.-Y., Brinton, C.G., Joe-Wong, C., Chiang, M.: Behavior-Based Grade Prediction for MOOCs Via Time Series Neural Networks. *IEEE Journal of Selected Topics in Signal Processing*, 11(5), pp. 716–728 (2017)
21. Yan, N., Tat-Sheung Au, O.: Online learning behavior analysis based on machine learning. *Asian Association of Open Universities Journal*, 14(2), pp. 97–106 (2019)
22. Rodríguez Armenta, C.E.: El Metacampus virtual. *Boletín Informativo CUDI* (2006)
23. Sistema de Universidad Virtual, Licenciatura en Tecnologías e Información, s/d 2020. [En línea]. Available: <http://www.udgvirtual.udg.mx/lti#>. [Último acceso: 30 marzo 2020] (2020)
24. Specht, D.F.: Probabilistic neural networks. *Neural Networks*, 3(1), pp. 109–118 (1990)
25. Masters, T.: *Advanced algorithms for neural networks*. Nueva York: Wiley (1995)
26. McLachlan, G. J.: *Discriminant Analysis and Statistical Pattern Recognition*. New York: Wiley (1992)
27. López-Martín, C.: Predictive accuracy comparison between neural networks and statistical regression for development effort of software projects. *Applied Soft Computing*, 27(1), pp. 434–449 (2015)