

UNA COMPARACIÓN DE REDES NEURONALES ARTIFICIALES EN EL PRONÓSTICO DE GRADUACIÓN

Jesús D Hernández Londoño
ICOM 6015 - Redes Neuronales Artificiales
Universidad de Puerto Rico Recinto Mayagüez

Resumen

Las bajas tasas de graduación es uno de los problemas más frecuentes en la educación superior. Las universidades buscan aumentar dichas tasas identificando las razones por las que los estudiantes no se gradúan. Por tanto, predecir la graduación, es decir, si el estudiante se gradúa o no, es de gran valor, puesto que, permite demostrar la eficacia de la institución ante los organismos de acreditación y el gobierno, y así proveer servicio necesario y oportuno para aumentar las posibilidades de que el estudiante complete su grado. El comportamiento prometedor de las Redes Neuronales Artificiales como clasificador, llevó en esta investigación a desarrollar redes neuronales artificiales multicapa teniendo en cuenta la variable Grade Point Average (GPA) para observar la incidencia de esta variable en los resultados predictivos de graduación.

1. Introducción

El análisis predictivo es un área del análisis de datos que pretende predecir tendencias y patrones de comportamiento. Actualmente, numerosas compañías como Facebook, Netflix y Amazon, están usando herramientas de análisis predictivo, para rastrear nuestros clics y llevarnos a elegir sus productos. La educación superior no se queda atrás, también están usando esta herramienta para mantener a los estudiantes inscritos abordando uno de los problemas más frecuentes en algunas universidades: las bajas tasas de graduación.

La incapacidad de los estudiantes para graduarse ha sido reflejo de la disminución de su rendimiento, debido a factores [6] como la formación académica previa, falta de interés o enfoque, selección incorrecta durante el proceso de admisión y distractores en el campus universitario [4] [5]. El Doctor Alberto Rodríguez, profesor de la Universidad Interamericana Recinto Ponce, [3] sugiere que esto lo causa la consejería académica insuficiente o inadecuada, los grupos grandes en los cursos, el ausentismo frecuente del estudiante, pocos cursos, entre otras. Adicionalmente, sugiere que establecer un sistema de alerta temprana, consejería académica compuesta por estudiantes, entre otras, ayudaría a aumentar la tasa de graduación.

Según el Informe Hechinger [1], “Las universidades están utilizando análisis predictivo para rastrear a los estudiantes en un esfuerzo por aumentar las tasas de graduación”, en este informe presentan el caso de la universidad Estatal de Georgia en el que el asesor académico de Keenan Robinson en su primer año lo orientó a elegir otro programa, dado que su plan de elegir enfermería era arriesgado porque no era probable que pasara el corte para dicho programa, aunque su Grade Point Average (GPA) era sólido el análisis predictivo detectó problemas.

De acuerdo con la Oficina de Planeación, Investigación y Mejoramiento Institucional (OPIMI) de la Universidad de Puerto Rico Recinto Mayagüez [2] la tasa de graduación, dentro del 150 % del tiempo del programa desde el año 1999 hasta 2010 ha oscilado entre 43.4 y 53.02 por ciento. En su informe del 2020, el 52,5 % de los estudiantes en el cohorte del 2011, habría completado su grado al cabo del 150 % del tiempo. Es decir, que aproximadamente solo la mitad de los estudiantes del Recinto de Mayagüez que comienzan la universidad obtienen sus títulos.

Por lo tanto, predecir la graduación de los estudiantes es de gran importancia para las universidades, dado que, permite demostrar la eficacia de la institución ante los organismos de acreditación y el gobierno [4].

Esta investigación presenta un análisis predictivo y propone el siguiente esquema:

- Predecir si los estudiantes subgraduados se gradúan o no, a través de Redes Neuronales Artificiales Multicapa teniendo en cuenta la variable Grade Point Average (GPA).
- Calcular la métrica Recall a partir de las predicciones obtenidas en las redes neuronales artificiales multicapa.
- Comparar la métrica Recall entre las redes neuronales artificiales multicapa.

2. Metodología

2.1. Datos

Los datos para esta investigación representan estudiantes subgraduados de la Universidad de Puerto Rico Recinto Mayagüez que completan sus estudios dentro del 150 % del tiempo del programa, en el periodo comprendido entre los años 1999 hasta 2010. La Universidad dispone una variedad de información sobre cada estudiante, sin embargo, para propósitos de esta investigación se considera la siguiente información:

Variable	Valor	Explicación
Year	Entero	Año de admisión
Faculty	Caracter	Facultad de admisión
Major	Factor	Programa de admisión
Apt Verbal	Entero	Extraído del Collage Board
Aprov Matem	Entero	Extraído del Collage Board
Apt Matem	Entero	Extraído del Collage Board
Aprov Espanol	Entero	Extraído del Collage Board
Aprov Ingles	Entero	Extraído del Collage Board
INGRESO FAMILIAR	Factor	Ingreso Familiar
EDUC PADRE	Factor	Educación del Padre (Grado Asociado o menos...)
EDUC MADRE	Factor	Educación de la Madre (Grado Asociado o menos...)
Gender	Caracter	Genero (Masculino, Femenino)
School Type	Factor	Tipo de Escuela (privada, pública, otra)
GPA 1ER ANO	Númérico	GPA del primer año en la universidad
GRAD	Factor	Graduación (si, no)
Rel Stud GPA	Númérico	Dificultad de la escuela
Rel School GPA	Númérico	Que tan bien las escuelas preparan a los estudiantes

Tabla 1: Descripción de variables para cada estudiante.

En total se analizarán 18,413 estudiantes, 14,731 se utilizaron para entrenar la red neuronal multicapa (entrenamiento 80 %) y los 3,682 restantes para prueba (prueba 20 %).

El software utilizado para el desarrollo de las redes neuronales fue Rstudio que es un Integrated Development Environment (IDE) para el lenguaje de programación R.

2.2. Arquitectura del Modelo

En esta investigación se utilizaron dos redes neuronales artificiales multicapa (ML), la primera red no tiene en cuenta la variable Grade Point Average (GPA), es decir, se eliminan temporalmente las variables GPA 1ER ANO y Rel School GPA, mientras que la segunda red tiene en cuenta la variable Grade Point Average (GPA), así que se añaden las variables GPA 1ER ANO y Rel School GPA.

La primera red se desarrolló [4] con tres capas. La primera capa (nivel de entrada) consta de 16 neuronas, una para cada variable, la tercera capa (nivel de salida), se compone de una sola neurona

Actual Positive	True Positive (TP)	False Negative (FN) Type I error
Actual Negative	False Positive (FP) Type II error	True Negative (TN)
	Predicted Positive	Predicted Negative

Figura 1: Matriz de confusión.

para denotar si el estudiante se gradúa o no. La segunda capa (nivel oculto) con 32 neuronas. La segunda red se desarrolló [4] igual que la primera red a diferencia de la segunda capa (nivel oculto) con 16 neuronas. Cada neurona se activa con la función Rectified Linear Unit (ReLU) puesto que arroja mejores resultados que la función tangente hiperbólica (TanH) que sugiere Karamouzis en [4], la neurona de salida se activa con la función Sigmoide. La tasa de aprendizaje que se usó fue por defecto 0.001 y la técnica de regularización usada más efectiva y más utilizada fue Dropout [8].

Por otro lado, el entrenamiento se implementó mediante el aprendizaje por lotes, es decir, se escoge un tamaño de muestra (o lote) que se presenta hacia adelante, luego se actualizan los pesos hacia atrás, al final este procedimiento constituye una época. Por lo general, el entrenamiento se lleva a cabo durante varias épocas. Para iniciar el entrenamiento, se utilizaron en los pesos, valores de una Distribucion Uniforme.

2.3. Métrica de Evaluación

Las métricas de evaluación son medidas que permiten calificar el rendimiento de un modelo de aprendizaje automático [7]. Una forma de representar el rendimiento de un modelo es la matriz de confusión que es una herramienta que permite visualizar de manera práctica el desempeño de un modelo.

La métrica que se describe a continuación se basa en el los TP y FN descritos en la figura 1.

2.3.1. Recall

El Recall, también conocido como la sensibilidad, es la proporción de casos positivos que fueron correctamente identificados por el modelo. Los casos positivos para esta investigación son los estudiantes que no se graduaron, dado que, se quiere minimizar el error de que un estudiante no se gradúa cuando el modelo dice que sí, permitiendo proveer servicio necesario para aumentar las posibilidades de que el estudiante complete su grado.

$$\frac{TP}{TP + FN} \quad (1)$$

Una interpretación informal del Recall formulado en la expresión 1, sería, cuantas veces el modelo clasifica correctamente los estudiantes que no se graduaron.

3. Resultados y Discusión

Se inicia con un preprocesamiento de datos que es la etapa en la que se limpian y reescalan los datos para continuar con la etapa de procesamiento, en la que se implementan los modelos. Siguiendo la sugerencia de Karamouzis [4] con 3 capas pero solo 4 neuronas en la capa oculta y 6000 épocas, se obtiene un Recall para la primera red de 0.34 y para la segunda red de 0.59. Luego de una revisión conjunta con Roberto Rivera Santiago Director de mi trabajo de grado se determina una alternativa, que reduce el tiempo de cómputo, utilizando dos capas ocultas de 16 neuronas para cada red con función de activación ReLU y 35 épocas con tamaño de lote de 600 estudiantes, obteniendo un Recall de 0.46 y 0.64 para la primera y segunda red respectivamente, la desventaja es mucho tiempo de cómputo para la sugerencia de Karamouzis, mientras que, si se considera 3 capas [4] y la capa oculta de la primera red tiene 32 neuronas y de la segunda red 16 neuronas con 35 épocas y tamaño de lote de 600 estudiantes se obtiene un Recall de 0.47, 0.64,

respectivamente, obteniendo una ligera mejora en el resultado como se muestra en la tabla 4 y 5.

		Valor Actual	
		No	Yes
Predicción	No	740	473
	Yes	820	1649

Tabla 2: Matriz de confusión para la predicción de graduación usando la primera red multicapa propuesta.

		Valor Actual	
		No	Yes
Predicción	No	985	3133
	Yes	575	1809

Tabla 3: Matriz de confusión para la predicción de graduación usando la Segunda red multicapa propuesta.

De la tabla 2 y 3 se puede obtener la probabilidad de que la red neuronal multicapa diga que el estudiante si se graduó dado que el estudiante realmente no se graduó. Para la primera y segunda red propuesta es de 0.53 y 0.36, respectivamente. Con lo anterior, se elige la segunda red neuronal multicapa propuesta porque tiene la menor probabilidad, indicando que al considerar la variable GPA el modelo minimiza el error de que un estudiante no se gradúa cuando el modelo dice que si.

Tipo de Red	Recall	Tiempo cómputo
ML Karamouzis	0.34	8681.98
ML Santiago	0.46	12.47
ML Propuesto	0.47	12.87

Tabla 4: Rendimiento primera red para datos de prueba.

Tipo de Red	Recall	Tiempo cómputo
ML Karamouzis	0.59	5284.99
ML Santiago	0.64	13.13
ML Propuesto	0.64	12.03

Tabla 5: Rendimiento segunda red para datos de prueba.

Como se esperaba la incidencia de la variable Grade Point Average (GPA) es bastante notoria como se muestra en la tabla 4 y 5. Es decir, que la red neuronal artificial multicapa propuesta que incluye la variable GPA es capaz de clasificar correctamente el 64 % de las veces los estudiantes que no se graduaron y el 36 % restante clasifica aleatoriamente.

Las figuras del 2 al 7, muestran el entrenamiento y validación (30 % para cada lote) de las redes neuronales artificiales multicapa y se observa un ligero sobreajuste para el ML Karamouzis y ML Santiago, incluso, para la red neuronal artificial multicapa propuesta en esta investigación, puesto que, la función de pérdida de validación (se muestra en verde) se incrementa mientras que el de entrenamiento (se muestra en rojo) decrece.

4. Conclusión

En general las redes neuronales artificiales multicapa trabajaron bien, a excepción del tiempo de cómputo que fue costoso para ML Karamouzis [4]. Se elige por simplicidad en la arquitectura, tiempo de cómputo y capacidad predictiva en la identificación de estudiantes en riesgo de no graduarse, la segunda red neuronal artificial multicapa propuesta. La segunda red neuronal multicapa propuesta que incluye la variable GPA es un modelo útil pero no excepcional en comparación con la primera red que no incluye la variable GPA.

5. Trabajo Futuro

En primera instancia se observo un ligero sobreajuste, así que, aumentar la base de datos debería perfeccionar el rendimiento de la red [4], también sería útil explorar validación cruzada como método para determinar cuándo detener la formación de la red y evitar el sobreajuste [9] mejorando el poder predictivo.

Determinar las variables que contribuyen a la identificación de estudiantes en riesgo de no graduarse y así proveer servicio necesario para aumentar las posibilidades de que el estudiante complete su grado.

Por último, aplicar la nueva arquitectura de aprendizaje de datos estructurados [10] TabNet desarrollado por Google que supera a variantes de árbol de decisiones y redes neuronales.

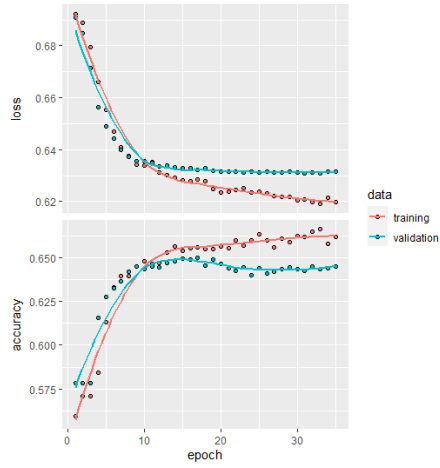


Figura 2: Función de pérdida y accuracy para datos de entrenamiento con ML Santiago primera red

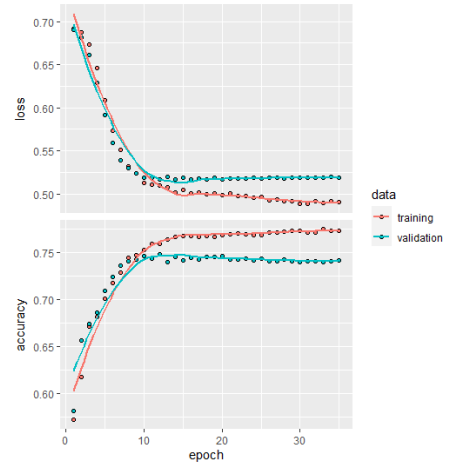


Figura 3: Función de pérdida y accuracy para datos de entrenamiento con ML Santiago segunda red

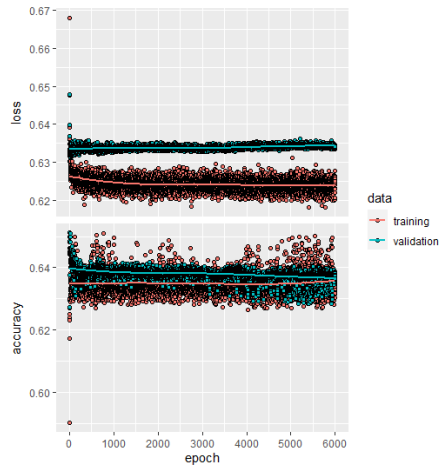


Figura 4: Función de pérdida y accuracy para datos de entrenamiento con ML Karamouzis primera red

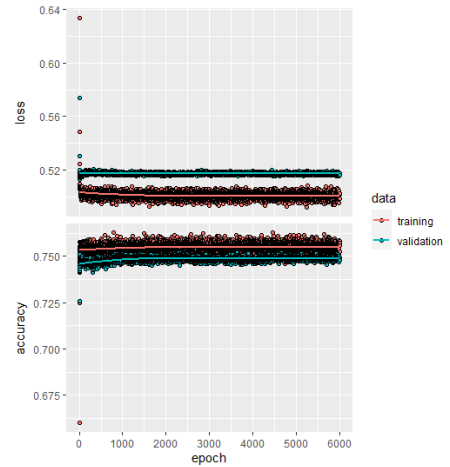


Figura 5: Función de pérdida y accuracy para datos de entrenamiento con ML Karamouzis segunda red

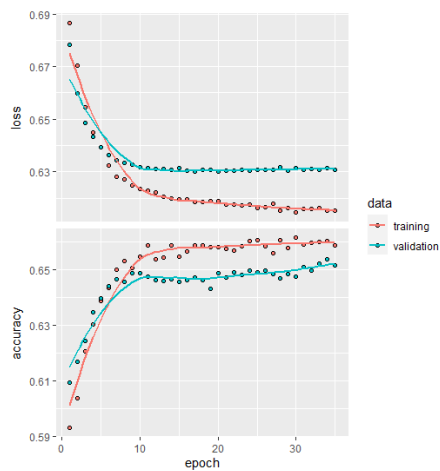


Figura 6: Función de pérdida y accuracy para datos de entrenamiento con ML Propuesta primera red

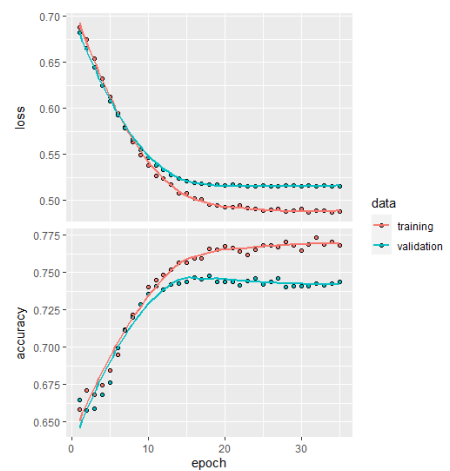


Figura 7: Función de pérdida y accuracy para datos de entrenamiento con ML Propuesta segunda red

6. Agradecimientos

Quiero agradecer a la Oficina de Planeación, Investigación y Mejoramiento Institucional (OPIMI) de la Universidad de Puerto Rico Recinto Mayagüez por facilitarme las bases de datos, a mi Director de trabajo de grado, Roberto Rivera Santiago y a mi hermano Jesus Andrés Hernandez, economista, por las sugerencias para el desarrollo de este investigación.

Referencias

- [1] J. Barshay, and S. Aslanian, "Colleges are using big data to track students in an effort to boost graduation rates, but it comes at a cost". *The Hechinger Report*. August 2019.
- [2] Oficina de Planeación, Investigación y Mejoramiento Institucional (OPIMI) [En Línea]. Disponible: <https://oiip.uprm.edu/tasas-de-graduacion-retencion-y-terminacion-a-tiempo/>
- [3] A. Rodriguez. 2010, Retención de Estudiantes [En Línea]. Disponible: <http://ponce.inter.edu/wp-content/uploads/documentos/retencion%20estudiantil/retencion%20%20estds%202010.pdf>
- [4] S.T. Karamouzis, and A. Vrettos, "An artificial neural network for predicting student graduation outcomes", in Proceedings of the World Congress on Engineering and Computer Science, in San Francisco, USA, 2008.
- [5] F.B. Ismail, and A. Marwan, Advance Intelligent Performance Prediction System.
- [6] J. S. Bassi, E. G. Dada, A. A. Hamidu and M. D. Elijah. "Students Graduation on Time Prediction Model Using Artificial Neural Network", *IOSR-JCE*, Vol. 21, Issue 3, PP 28-35 , May - June 2019.
- [7] E. C. Ploutz, "Machine Learning Applications in Graduation Prediction at the University of Nevada, Las Vegas".B.S. Theses, University of Nevada, Las Vegas, NV, USA, 2018. <http://dx.doi.org/10.34917/13568668>
- [8] N. Srivastava, G Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", *JMLR*, Vol. 15, PP 1929-1958, June 2014.
- [9] M. Stone, "Asymptotics for and against cross-validation", *Biometrika*, Vol. 64, PP 29-35, 1977.
- [10] S. O. Arik, and T. Pfister, "TabNet: Attentive Interpretable Tabular Learning", *arXiv, [cs.LG]*, V. 4, 2020. <https://arxiv.org/pdf/1908.07442.pdf>