

Vea discusiones, estadísticas y perfiles de autor para esta publicación en: <https://www.researchgate.net/publication/344217296>

Evaluación del rendimiento de los estudiantes mediante algoritmos de aprendizaje automático

Artículo · Septiembre 2020

CITAS

0

LEE

900

1 autor:



[Samah Fakhri Aziz](#)

Universidad de Al-Hamdaniya

4 PUBLICACIONES 1 CITA

VER EL PERFIL

Evaluación del desempeño de los estudiantes usando la máquina
Algoritmos de aprendizaje

Samah Fakhri Aziz

Departamento de Ciencias de la Computación
Universidad de AL-Hamdaniya
Mosul, Irak

Recibido: 18/5/2020 ; Aceptado: 5/7/2020

Resumen :

Predecir el desempeño de los estudiantes es muy importante para el éxito de cualquier proceso educativo. Aprovechamiento de métodos de minería de datos y máquina. aprender a predecir su desempeño en base a los datos disponibles en las escuelas y los registros de los estudiantes pueden explicar su comportamiento, el impacto de cada factor sobre el progreso del proceso educativo de los estudiantes, la relación de la etapa de edad y seguimiento de padres y días de ausencia. Esta artículo analiza la posibilidad de aprovechar el aprendizaje automático algoritmos para predecir el desempeño de los estudiantes y determinar la importancia de cada factor con ese desempeño y Comparando el desempeño de algoritmos de aprendizaje automático (GBDT-RFDT-Deeplearning) en la exploración datos educativos.

Palabras clave: rendimiento de los estudiantes, GBDT, RFDT, aprendizaje profundo, predicción.

Dominar el desempeño del alumno, utilizando las herramientas de la publicidad

Resumen:

Conclusión: el bostezo es la realización de la prolongación, lo que es una ventaja para el éxito de cualquier dependencia de elevación. Poder extraer los extractos de los tercios y anunciar su actuación según los tres disponibles en las manos y los registros del estudiante, es decir, el desplazamiento de sus hilos y la influencia de cada monarca en la escala de la supremacía del estudiante. y la relación que es incluso la viuda y la viuda. Este trabajo discute el agotamiento de la actuación de los algoritmos publicitarios para tensar la actuación de la prolongación y la distorsión de los deseos de cada monarca por esta actuación, y el agotamiento de la actuación de los algoritmos publicitarios (GBDT-GBDT) (Deeplearning-RFDT) en Exploring the Lexical Threes.

Liposucciones - estiramientos, bostezos.

I. INTRODUCCIÓN

El descubrimiento de conocimiento a partir de grandes bases de datos se conoce como minería de datos. El propósito de este proceso es extraer información oculta o patrones repetitivos que pueden ser útiles en muchas ciencias. La minería de datos se ha utilizado en muchos campos, como la medicina, la economía, el marketing y la administración de empresas.

[1]. Los datos del sistema educativo pueden ser de utilidad para extraer información y así obtener conocimientos que ayuden a la toma de decisiones y al mejoramiento del proceso educativo [2].

La exploración en los expedientes de los alumnos puede ser útil para conocer las razones del bajo rendimiento de algunos alumnos, identificar las causas de determinados comportamientos, incluso las causas de algunas enfermedades que les afectan y la predicción temprana del nivel de los alumnos en base a sus datos para ayudar. ellos [3]. Los algoritmos de aprendizaje automático son una de las herramientas utilizadas en la minería de datos y la extracción de conocimiento y han demostrado su eficacia en muchos

campos. Para predecir el rendimiento de los estudiantes, este artículo sugiere el uso de algoritmos de árboles de decisión avanzados (GBDT RFDT) y comparar su rendimiento con el aprendizaje profundo.

II. ALGORITMOS DE MINERÍA DE DATOS

Los algoritmos de minería de datos ayudan a encontrar los factores que afectan el desempeño de los estudiantes y pronostican su desempeño. Los algoritmos de minería se clasifican en términos de cómo funcionan, como clasificación, agrupamiento, aprendizaje automático, regla de asociación, etc. En este trabajo, los algoritmos de aprendizaje automático se utilizan para predecir el rendimiento de los estudiantes.

A. Bosque aleatorio (RF)

Random Forest (RF) es un método de aprendizaje de conjunto utilizado para la clasificación, la regresión que funciona mediante la creación de muchos árboles de decisión en el momento del entrenamiento. Los bosques aleatorios corrigen el hábito de los árboles de decisión de ajustarse demasiado al conjunto de entrenamiento. El entrenamiento para bosques aleatorios aplica la técnica general de embolsado a los aprendices de árboles [5].

B. Aumento de gradiente (GB)

El refuerzo de gradiente (GB) es un método para convertir a los alumnos débiles en alumnos fuertes al aumentar el peso de las propiedades que son difíciles de clasificar. Después de crear el primer árbol, se cambian los pesos y se crea un subárbol a partir de ellos.

Cada árbol es una versión modificada del conjunto de datos original, y los resultados de esos árboles se toman para lograr resultados más precisos con fines de clasificación y pronóstico[6][7].

C) Aprendizaje profundo

El aprendizaje profundo es un grupo de unidades de E/S conectadas y cada conexión tiene un peso. Estas redes aprenden ajustando pesos entrenándolos en modelos que conocen sus resultados. Esta red tiene la capacidad de manejar valores continuos de entradas y salidas y definir patrones. Es adecuado para propósitos de pronóstico y clasificación.

tercero METODOLOGÍAS SUGERIDAS

Los algoritmos propuestos en esta investigación se aplican al conjunto de datos recopilados de los organismos educativos. El número de estudiantes a los que se les aplicó el estudio fue de 450 instancias y 17 atributos. Las características tomadas para predecir el rendimiento de los estudiantes se muestran en la siguiente tabla.

TABLA 1: atributo alumno tomado

Sin atributo	Variable	Posible Valores
1 edad	numéricos	17-28
2 Género	binario MF	
3 ID de etapa	nominal	primero, segundo, Tercero cuarto
4 Relación nominal Padre, Madre		
5 Manos levantadas numéricas 0-100		
6 StudentAbsenceDays numérico 0-25		
7 Logro académico de la madre numérico 1-2-3-4		
8 Logros académicos del padre numérico 1-2-3-4		
9 El trabajo del padre	nominal	Gubernamental, comercial, residencial
10 trabajo de madre	nominal	Gubernamental, comercial, residencial
11 Estado de los padres	casado nominal,	divorciado, muerto
12 Núm. de fracasos de clase pasados	numérico	0-5
13 Tiempo de viaje	numérico	1-2-3
14 tiempo de estudio	numérico	1-2-3
15 Estado de salud	binario	bueno, malo
16 Internet	numérico	1-2-3
17 Clase	HML nominal	(objetivo)

Los niveles de desempeño de los estudiantes se dividieron en tres niveles (alto, medio y bajo). Según sus calificaciones y sus niveles, se distribuyeron como se muestra en la Figura (1).

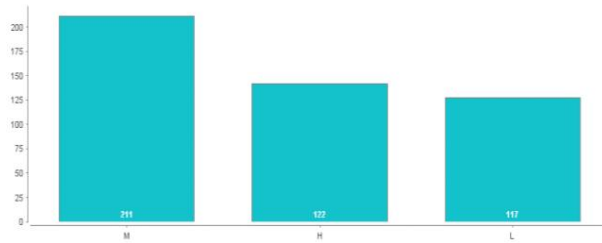


Figura (1): Estudiante

niveles de rendimiento

Los datos se procesan previamente y se comprueban los valores perdidos. Se tuvieron en cuenta varios factores, incluido el rendimiento de los algoritmos propuestos en términos de precisión en el pronóstico, el tiempo de ejecución del entrenamiento, el pronóstico y el impacto de cada función en los resultados.

La metodología utilizada en esta investigación y sus etapas se muestra a continuación:

1. Primero, los algoritmos de aprendizaje automático que aumentan el gradiente del árbol de decisión y el árbol de decisión de bosque aleatorio y la red neuronal profunda se entrenaron en conjuntos de datos educativos y Se obtuvieron modelos.
2. Los modelos entrenados obtenidos en la primera etapa se aplican a los datos de prueba, ya que el conjunto de datos se dividió en un 80 % de grupo de entrenamiento y un 20 % de grupo de prueba.
3. Los resultados de cada algoritmo se analizan y comparan en términos de precisión, tiempo de predicción y tasa de error.

Las etapas de predicción para esta investigación se resumen a continuación.

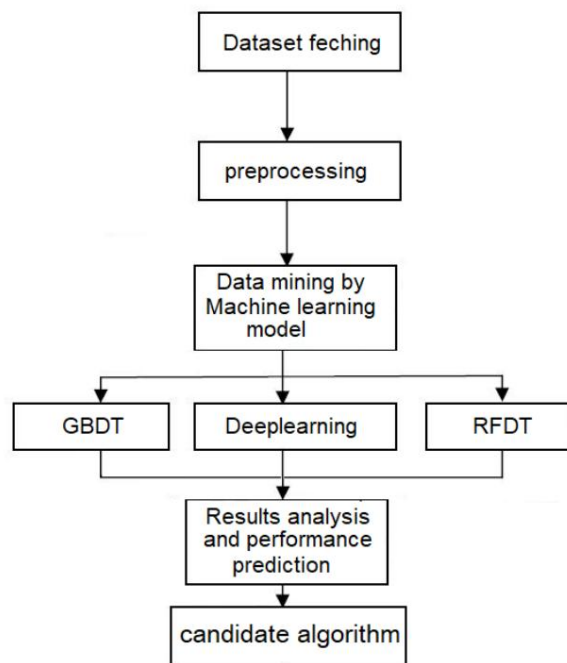


Fig (2): Marco de Trabajo Propuesto

IV. RESULTADOS Y DISCUSIONES

En este trabajo se seleccionaron varios clasificadores y se realizó un análisis comparativo de su desempeño utilizando la herramienta Rapidminer. El conjunto de datos educativos se prepara previamente y luego se proporciona para GBDT, RFDT y Deeplearning. Estos algoritmos fueron entrenados y probados y los resultados finales se muestran en la siguiente tabla:

TABLA 2: precisión de la predicción

Precisión del modelo		Estándar Desviación
Profundo Aprendiendo	0.782	0.0373
Aleatorio bosque	0.761	0.035
Degradado Impulsado Árboles	0.716	0.0454

La Tabla 2 muestra la precisión de la predicción para cada algoritmo utilizado en este documento.

Para mostrar los errores estadísticos de cada modelo se utilizó Confusion Matrix y los resultados fueron los siguientes:

TABLA 3: Matriz de confusión de RFDT

	MLH		
M 34		2	14
L	4	33	1
H 16		0	32

TABLA 4: Matriz de confusión de GBDT

	MLH		
M 39 4 9			
L	4	34 1	
H 17		0 79	

TABLA 5: Matriz de confusión de Deeplearning

	MLH		
METRO 41			1
L 6		7	1
H 9		24 0	21

Una métrica importante para realizar algoritmos de aprendizaje automático es una curva característica operativa del receptor o curva ROC. Esta escala ofrece una vista visual de la capacidad de separación de los algoritmos de clasificación. En este trabajo, cada desempeño de cada algoritmo se expresa en tres curvas, donde el símbolo de la curva que representa el nivel alto del estudiante con (H) y la mediana con (M) y bajo con (L). Esta escala da una visión visual de la capacidad de separación de los algoritmos de clasificación. En esta investigación, cada rendimiento de cada algoritmo se expresa en tres curvas, donde el símbolo de la curva que representa

nivel de estudiante alto con (H) y medio con (M) y bajo con (L) como se muestra a continuación:

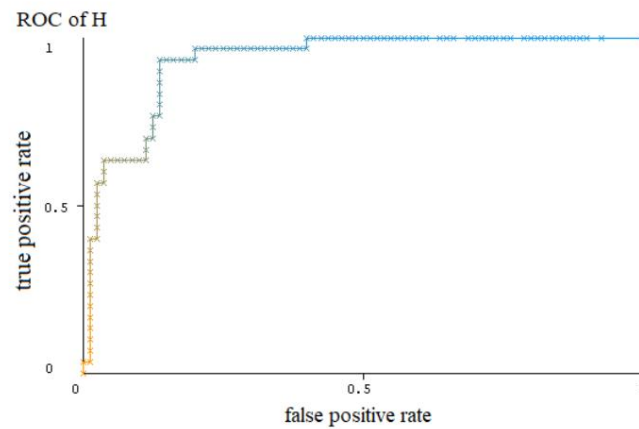


Fig (3): Curva ROC de H usando Random Forest

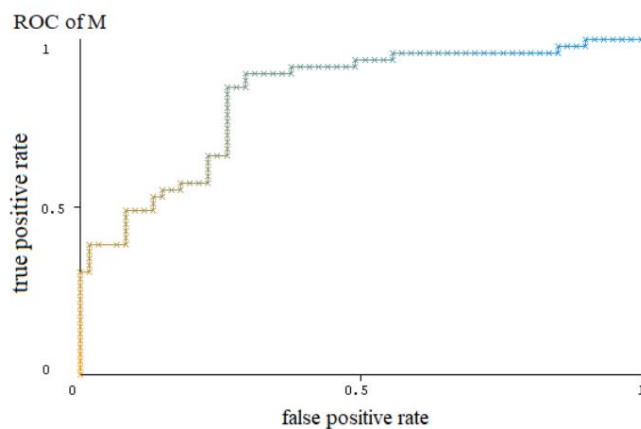


Fig (4): Curva ROC de M usando Random Forest

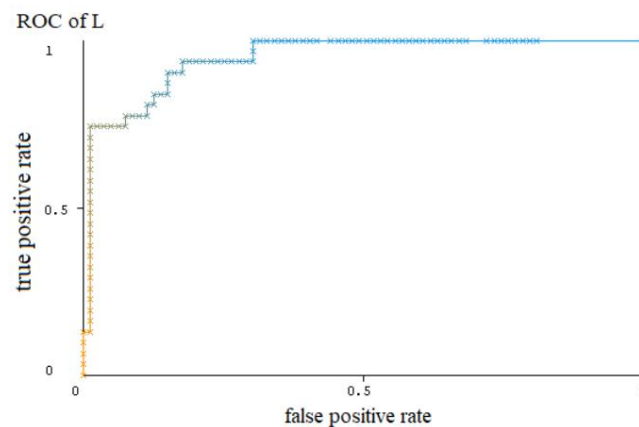


Fig (5): Curva ROC de L usando Random Forest

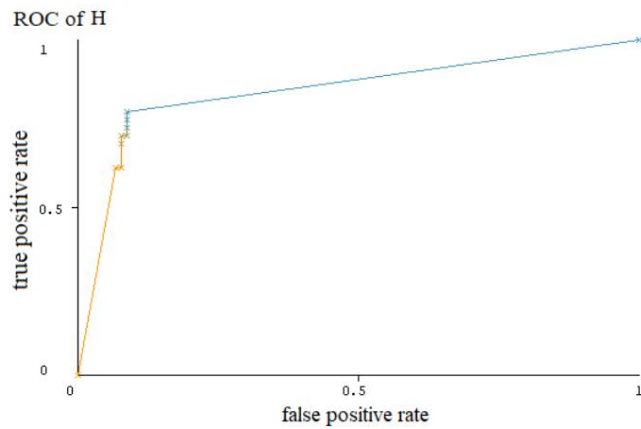


Fig (6): Curva ROC de H usando árboles potenciados por gradiente

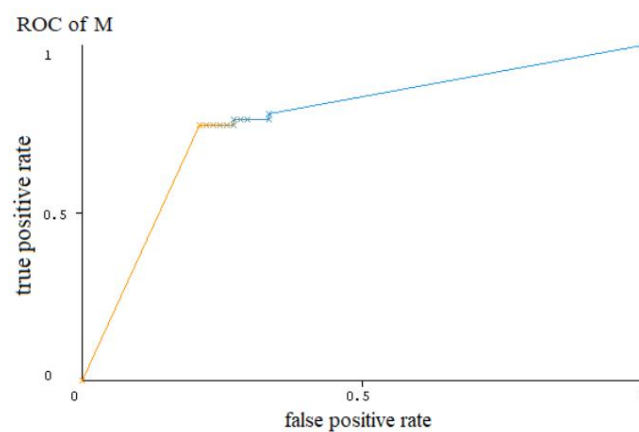


Fig (7): Curva ROC de M usando Gradient Boost Trees

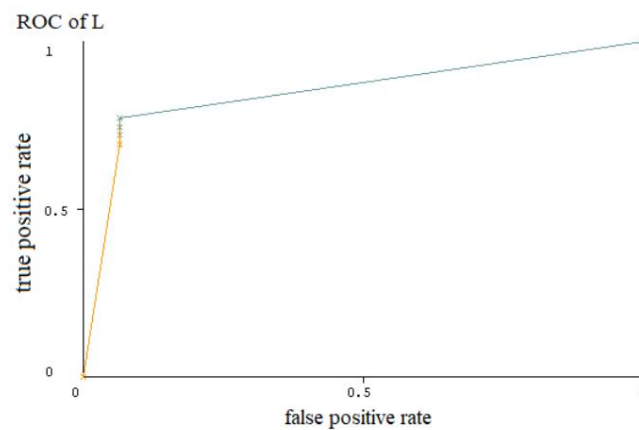


Fig (8): Curva ROC de L usando Gradient Boost Trees

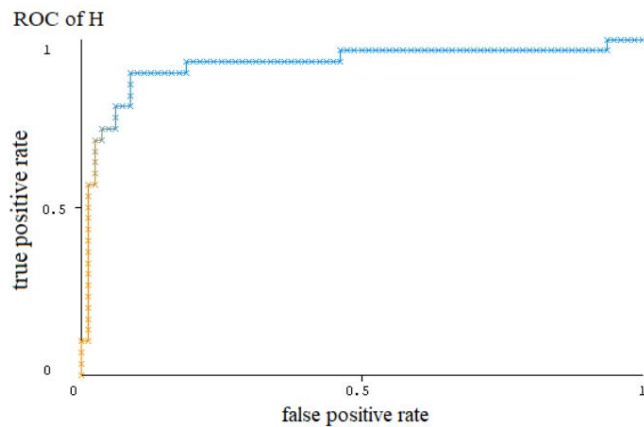


Fig (9): Curva ROC de H usando Deep Learning

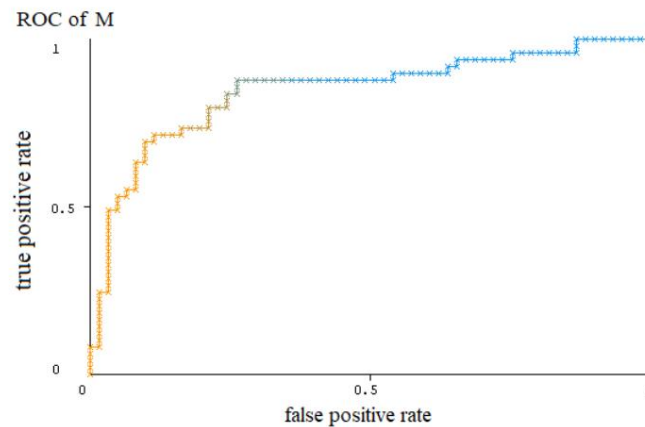


Fig (10): Curva ROC de M usando Deep Learning

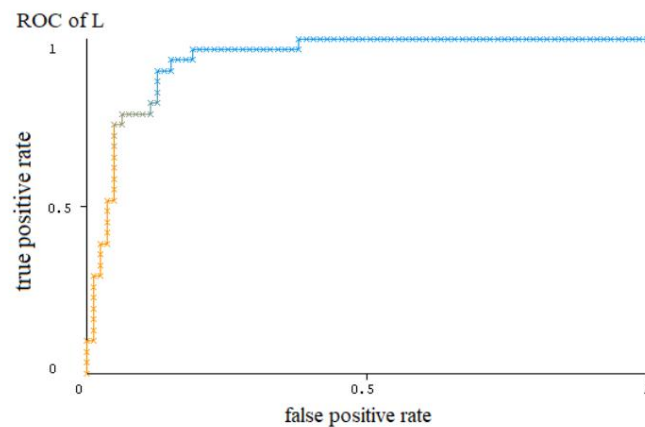


Fig (11): Curva ROC de L usando Deep Learning

Como se muestra en las Figuras 3, 5, 9, 10 y 11 es la que tiene mayor susceptibilidad a la separación ya que las curvas de las Figuras 9, 10, 11 pertenecen a Deep Learning y esto demuestra que son apropiadas para una clasificación de todas niveles de desempeño de los estudiantes.

CONCLUSIÓN

La minería de datos en el sistema educativo es extremadamente importante para analizar y anticipar el rendimiento académico de los estudiantes al observar diferentes factores de rendimiento. Este estudio brindará una solución para pre-evaluar el desempeño de los estudiantes, lo que a su vez ayuda a desarrollar el desempeño de los estudiantes y cuidarlo de manera oportuna en la dirección correcta. En este estudio se ha comparado el rendimiento de los algoritmos de aprendizaje automático con fines de minería de datos en instituciones educativas y se ha demostrado que estos algoritmos son prometedores en el campo de la previsión del rendimiento de los estudiantes y el Deep Learning es un algoritmo apropiado para este tipo de conjunto de datos

REFERENCIAS

- [1] Sultana, J., Rani, MU y Farquard, MAH (2019). Predicción del rendimiento de los estudiantes utilizando métodos de aprendizaje profundo y minería de datos. *Revista internacional de tecnología e ingeniería recientes (IJRTE)*, 8 (1S4), 2277-3878.
- [2] Scheuer, O. y McLaren, BM (2012). Minería de datos educativos. *Enciclopedia de las ciencias del aprendizaje*, Springer, 1075-1079., Doi: [10.1007/978-1-4419-1428-6_618](https://doi.org/10.1007/978-1-4419-1428-6_618).
- [3] Romero, C. y Ventura, S. (2010). Minería de datos educativos: una revisión del estado del arte. *Transacciones IEEE en sistemas, Man y Cybernetics, Parte C (Aplicaciones y revisiones)*, 40(6), 601-618.
- [4] Zaker, NA, Alsaleem, N. y Kashmoola, MA (2019). Solución de modelos multiagente para lograr EMC en sistemas de telecomunicaciones inalámbricas. En *Actas - 2018 1ra Conferencia Internacional Anual sobre Información y Ciencias, AiCIS 2018* (págs. 311–314). Instituto de Ingenieros Eléctricos y Electrónicos Inc. <https://doi.org/10.1109/AiCIS.2018.00061>.
- [5] Ho, TK (1995). Bosques de decisión aleatoria. En *Actas de la 3ra conferencia internacional sobre análisis de documentos y reconocimiento*, IEEE, 1, 278-282. doi: [10.1109/ICDAR.1995.598994](https://doi.org/10.1109/ICDAR.1995.598994)
Archivado desde [el original](#) (PDF) el 17 de abril de 2016.
- [6] Friedman, JH y Meulman, JJ (2003). Árboles de regresión aditivos múltiples con aplicación en epidemiología. *Estadísticas en medicina*, 22(9), 1365-1381. doi: 10.1002/sim.1501.
- [7] Elith, J. y Leathwick, J. (2017). Árboles de regresión potenciados para modelado ecológico. *R Documentación*. Disponible en línea: <https://cran.r-project.org/web/packages/dismo/viñetas/brt.pdf> (consultado el 12 de junio de 2011).
- [8] Baradwaj, BK y Pal, S. (2012). Minería de datos educativos para analizar el desempeño de los estudiantes. *IJACSA*, 2(6), 63-69. *arXiv preimpresión arXiv:1201.3417*.
- [9] Asif, R., Merceron, A. y Pathan, MK (2014). Predicción del rendimiento académico de los estudiantes a nivel de grado: un estudio de caso. *Revista internacional de sistemas y aplicaciones inteligentes*, 7(1), 49. doi: 10.5815/ijisa.2015.01.05.
- [10] Lemeshko, O., Yevdokymenko, M. y Alsaleem, NYA (2018). Desarrollo del modelo tensorial de enrutamiento QoE multicamino en una red de infocomunicaciones proporcionando la calificación de calidad requerida. *Diario de Europa del Este de Tecnologías Empresariales*, 5(2), 40–46. <https://doi.org/10.15587/1729-4061.2018.141989>.
- [11] Osmanbegovic, E. y Suljic, M. (2012). Enfoque de minería de datos para predecir el rendimiento de los estudiantes. *Revisión económica: Revista de Economía y Negocios*, 10(1), 3-12. ISSN: 1512-8962.
- [12] Alsaleem, NYA, Kashmoola, MA y Moskalets, M. (2018). Análisis de la eficiencia del acceso espaciotemporal en los sistemas de comunicaciones móviles basados en un arreglo de antenas. *Diario de Europa del Este de Tecnologías Empresariales*, 6(9–96), 38–47. <https://doi.org/10.15587/1729-4061.2018.150921>.