

Documento de investigación original

Predicción del Rendimiento Académico de los Estudiantes en la Universidad

Uso de clasificadores de metaárboles de decisión

¹Shanthini, A., 2G. Vinodhini y 2R.M. chandrasekaran¹Departamento de Tecnología de la Información, Universidad SRM, Chennai, India²Departamento de Informática e Ingeniería, Universidad de Annamalai, Chidambaram, India

Historial del artículo

Recibido: 05-05-2017

Revisado: 05-10-2017

Aceptado: 19-02-2018

Autor para

correspondencia: Shanthini, A.

Departamento de Tecnología

de la Información, Universidad

SRM, Chennai, India Correo

electrónico: shanthini.aus@gmail.com

Resumen: La predicción del rendimiento de los estudiantes es un área de preocupación para las instituciones educativas. En el sistema de aprendizaje de nivel universitario, el método o la regla adoptada para identificar a los candidatos que aprueban o suspenden difiere según varios factores, como el curso, el departamento de estudio, etc. Predecir el resultado de un alumno en un curso es un tema que se ha abordado recientemente utilizando técnicas de aprendizaje automático. El objetivo de este trabajo es encontrar una manera de predecir el rendimiento académico de un estudiante en la Universidad utilizando el enfoque de aprendizaje automático. Esto se hace utilizando los registros anteriores del estudiante en lugar de aplicar fórmulas dependientes del curso para predecir la calificación final del estudiante. En este trabajo, se utilizan técnicas de clasificación de metaárboles de decisión basadas en cuatro algoritmos de aprendizaje representativos, a saber, Adaboost, Bagging, Dagging y Grading, para construir diferentes árboles de decisión. REPTree se utiliza como método de árbol de decisiones para el metaaprendizaje. Estos cuatro métodos de metaaprendizaje se han comparado por separado con respecto a los conjuntos de entrenamiento y prueba. Adaboost resulta ser el mejor clasificador de metadecisiones para predecir el resultado del alumno en función de las notas obtenidas en el semestre.

Palabras clave: árboles de decisión, metaclasificadores, REPTree, aprendizaje automático

Introducción

Las universidades operan en entornos viables muy enérgicos y efectivos. Un volumen masivo de datos con referencia a los estudiantes está disponible en formato digital.

Sin embargo, la conversión efectiva de estos voluminosos datos en conocimiento para la toma de decisiones es un gran problema. Predecir el desempeño de un estudiante es uno de esos desafíos que enfrenta el sector educativo (Asogbon et al., 2016). En los últimos años, muchos trabajos de investigación se han centrado en técnicas de minería de datos en instituciones de educación superior para mejorar el método de aprendizaje. El trabajo existente en la predicción de los estudiantes

El rendimiento incluye analizar los datos de matriculación de los estudiantes para evitar la deserción, predecir la detención de los estudiantes en una etapa temprana y analizar la calidad y el uso de los materiales de aprendizaje. El desarrollo de un sistema automatizado para esto ayudará a los educadores a monitorear el desempeño de sus estudiantes. La sección 2 presenta los logros (Buldu y Ucgun, 2010; Delen, 2010; Marquez-Vera et al., 2016) y los estudiantes para mejorar sus habilidades de aprendizaje. El sistema automatizado también ayudará al personal administrativo a mejorar la

desempeño de las instituciones. Así, la aplicación de técnicas de minería de datos puede enfocarse en aplicaciones particulares de un sistema automatizado (Amrieh et al., 2016; Chen and Bai, 2010; Hien and Haddawy, 2007; Nespereira et al., 2015).

Un conjunto de diferentes algoritmos de aprendizaje automático es un método efectivo que se utiliza para adquirir un alto nivel de precisión predictiva. Sin embargo, dichas mejoras se basan en los conjuntos existentes. Si todas las instancias del conjunto se comportan casi de la misma manera, se logra un poco más combinando sus predicciones. Los árboles de decisión se adoptan mejor para los métodos de conjunto, ya que son rápidos y estables (Buldu y Ucgun, 2010). El objetivo principal de este trabajo es comparar el rendimiento de varios algoritmos de árboles de metadecisiones para predecir el rendimiento de los estudiantes tanto en conjuntos de entrenamiento como en conjuntos de prueba. El resto de este trabajo se estructura de la siguiente manera. La sección 2 presenta el fondo. La sección 3 describe los modelos de datos utilizados. La sección 4 presenta los métodos adoptados y la sección 5 discute los resultados seguidos por la sección 6 con las conclusiones y posibilidades de investigación futuras.

Fondo

La predicción del rendimiento de un estudiante se ha estudiado previamente en la investigación de minería de datos educativos en el contexto de la deserción estudiantil. Wolff et al. (2014) exploró la eficacia de los métodos de modelado predictivo para identificar a los estudiantes que se beneficiarán más de las intervenciones de los tutores en el aprendizaje a distancia. Los estudiantes y el tutor no se encontrarán cara a cara en el caso de la educación a distancia.

Los métodos analizados para el aprendizaje a distancia incluyeron clasificación de árboles de decisión, máquina de vectores de soporte, autómata de hipótesis unarias generales, redes bayesianas y regresión lineal y logística. Romero et al. (2013) investigó cómo la precisión de la predicción se vio afectada por factores como la selección de instancias y atributos, el uso de algoritmos de clasificación y la fecha en que se recopilaban los datos. Se desarrolló un nuevo módulo de Moodle para la recolección de indicadores del foro. Utilizando este Moodle, se llevaron a cabo diferentes experimentos con datos reales de 114 estudiantes universitarios de un curso de primer año de informática. Los resultados obtenidos demostraron su eficacia tanto en términos de predicción final al final del curso como de predicción temprana antes del final del curso.

Márquez-Vera et al. (2016) propusieron una técnica y un algoritmo de clasificación para construir un modelo de predicción del abandono escolar lo antes posible. El conjunto de datos utilizado para su investigación se obtuvo de 419 estudiantes de secundaria en México. Se llevaron a cabo varios experimentos para predecir la deserción en diferentes niveles del curso. Se compararon algunos algoritmos de clasificación bien conocidos clásicos y desequilibrados con su algoritmo propuesto para dar el mejor indicador de abandonos. Los resultados obtenidos en su trabajo de investigación mostraron que el algoritmo ideado por ellos fue lo suficientemente efectivo para predecir la deserción de los estudiantes dentro de las primeras 4-6 semanas del curso. El algoritmo se puede utilizar como un sistema de alerta temprana. Ramesh et al. (2013) adoptaron una metodología experimental para generar una base de datos construida a partir de fuentes primarias y secundarias. Los resultados obtenidos por este trabajo revelan que la ocupación de los padres juega un papel importante en la predicción de la calificación de los estudiantes, mientras que el tipo de escuela no influyó en los resultados de los estudiantes. Dichos hallazgos pueden ayudar a las instituciones a identificar a los estudiantes débiles en riesgo y concentrarse en brindarles capacitación adicional.

Zhang et al. (2015), con el fin de mejorar el nivel académico en la etapa de pregrado y lograr una mejor calificación de tesis de graduación, investigó las relaciones ocultas entre cursos y calificaciones de tesis de graduación y empleó la máquina de vectores de soporte para construir un modelo de predicción para predecir las calificaciones de tesis de graduación. de estudiantes universitarios. También se construyeron otros modelos de predicción (Neural Network, Decision Tree y Naïve Bayes) para predecir las calificaciones de las tesis de graduación de los estudiantes universitarios, pero el resultado mostró que la Máquina de vectores de soporte (SVM) se desempeñó mejor en este caso de estudio. Strecht et al. (2015) abordó el problema de

predecir el éxito o el fracaso de un estudiante en un curso o programa utilizando técnicas de minería de datos. Evaluaron algunos de los algoritmos de clasificación y regresión más populares en este problema. Abordaron dos problemas en particular: predicción de aprobación/reprobación y predicción de calificación. Los algoritmos con mejores resultados globales en clasificación fueron árboles de decisión y SVM mientras que en regresión fueron SVM, Random Forest y AdaBoost (Illanas Vila et al., 2013). Arsad et al. (2013) utilizaron un modelo de red neuronal artificial (ANN) para predecir el rendimiento académico de los estudiantes de ingeniería que cursaban una licenciatura. El estudio toma el Punto de Calificación (GP) obtenido por los estudiantes en las materias fundamentales como entradas sin considerar su origen demográfico, mientras que toma el Promedio de Calificación Acumulativo (CGPA) como salida. Schalk et al. (2011) construyeron un sistema predictivo basado en aprendizaje automático para determinar qué estudiantes estaban en riesgo de reprobar cursos introductorios de matemáticas y física. El sistema utilizó la técnica Random Forest para modelar datos provenientes de años anteriores del SAT. Si bien sus resultados fueron buenos, el método diseñado no pretendía mantenerse en el tiempo ni hacer predicciones progresivas basadas en información incremental.

En resumen, se han investigado diversas investigaciones para resolver problemas educativos utilizando técnicas de minería de datos. Sin embargo, muy pocas investigaciones han arrojado luz sobre el comportamiento de los estudiantes durante el proceso de aprendizaje y su impacto en el éxito académico de los estudiantes. Esta propuesta de la investigación se centrará en el impacto del sistema académico en el rendimiento de los estudiantes. El rendimiento del modelo predictivo del estudiante se evaluó mediante un conjunto de clasificadores, a saber, algoritmos ANN, Naïve Bayesian y Decision Tree. Además, aplicamos métodos de conjunto para mejorar el rendimiento de estos clasificadores. El conocimiento extraído ayudará a las escuelas a mejorar el desempeño académico de los estudiantes y ayudará a los administradores a mejorar los sistemas de aprendizaje. Este trabajo también se centra, en particular, en el rendimiento de los alumnos para predecir si el alumno aprobará o reprobará al final de la carrera para diferenciar entre los alumnos fuertes y los débiles. El estudio de investigación actual difiere de otros trabajos al limitar las variables en la predicción del rendimiento solo a las calificaciones; no se utilizaron datos demográficos ni socioeconómicos. Se necesita un conjunto de entrenamiento para construir un modelo y otro conjunto de prueba para evaluarlo, lo que permite medir qué tan bien se pueden generalizar los hallazgos.

Metodología

En este estudio, presentamos un modelo de interpretación para estudiantes que utilizan métodos de conjunto. Un método de conjunto es un enfoque de aprendizaje que combina múltiples modelos para resolver un problema. A diferencia de los enfoques de aprendizaje tradicionales que entrenan los datos mediante un modelo de aprendizaje, los métodos de conjunto intentan entrenar los datos con un

y luego combínelos para votar sus resultados.

Las predicciones hechas por conjuntos suelen ser más precisas que las predicciones hechas por un solo modelo. En la figura 1 se muestra el diseño metodológico de este trabajo de investigación.

Los siguientes son los pasos involucrados en el diseño de la metodología:

- Recopile datos e identifique las características de los conjuntos de datos . • Para un conjunto de datos, desarrolle los siguientes modelos de predicción utilizando el conjunto de datos de entrenamiento respectivo.

- Embolsado. • Impulso. • Dagging
- y • Clasificación.

- Predecir las marcas para todos los modelos mencionados anteriormente • Comparar los resultados de la predicción con los resultados reales

El método REPTree se utiliza como método de predicción base para todos los métodos de conjunto empleados. El algoritmo REPTree se basa en el principio de calcular la ganancia de información con entropía y reducir el error que surge de la varianza. La complejidad del modelo de árbol de decisión se reduce mediante el "método de poda de error reducido" y se reduce el error que surge de la varianza.

impulsar

En Boosting, se selecciona un subconjunto aleatorio de muestras de entrenamiento d1 sin reemplazo del conjunto de entrenamiento D para entrenar a un alumno débil C1. Para seleccionar un segundo subconjunto de entrenamiento aleatorio d2 sin reemplazo del conjunto de entrenamiento, agregue el 50 por ciento de las muestras que previamente se clasificaron incorrectamente para entrenar a un alumno débil C2. Encuentra el

muestras de entrenamiento d3 en el conjunto de entrenamiento D en el que C1 y C2 no están de acuerdo para entrenar a un tercer alumno débil C3. Combinar a todos los alumnos débiles a través de la votación por mayoría (Petkovic et al., 2012).

Harpillera

Bagging representa a Bootstrap AGGREG en ING. Los pasos involucrados en el embolsado son los siguientes:

- Generar n muestras de entrenamiento de arranque diferentes • Entrenar el algoritmo en cada muestra de arranque por separado • Promediar las predicciones al final

Dagging

El clasificador Dagging Meta crea una serie de pliegues estratificados en conexos a partir de los datos y alimenta cada fragmento de datos a una copia del clasificador base proporcionado. Con la votación por mayoría, las predicciones se realizan en clasificadores básicos que se acumulan en el clasificador Vote Meta. Se identifican varias instancias de entrenamiento útiles para clasificadores base que son cuadráticos o peores en el comportamiento del tiempo (Sorour et al., 2015).

calificación

La idea subyacente de calificar es predecir si la predicción para un ejemplo en particular es correcta o no para cada uno de los algoritmos de aprendizaje originales. Por lo tanto, se entrena un clasificador para cada algoritmo de aprendizaje original con el conjunto de entrenamiento de ejemplos originales.

Estos ejemplos originales tienen etiquetas de clase que codifican si la predicción de los alumnos fue correcta en este ejemplo en particular (Yoo y Kim, 2014).

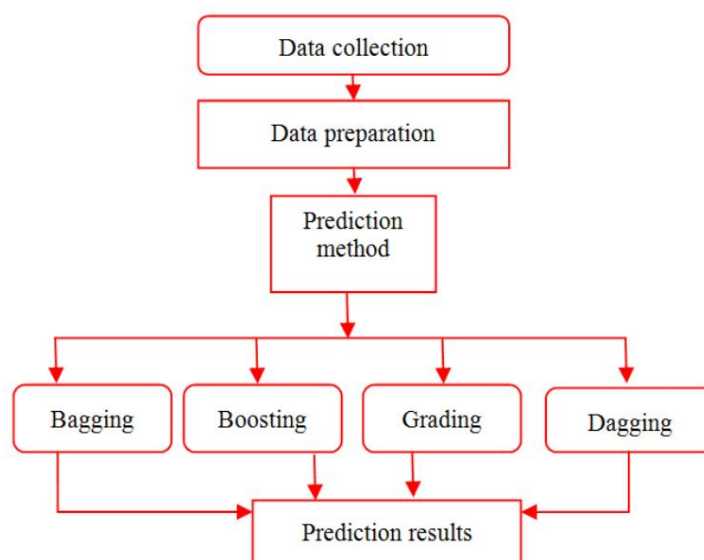


Fig. 1: Diseño de la metodología

Resultados

En este estudio, utilizamos las calificaciones obtenidas por los estudiantes de informática e ingeniería en dos semestres. Cada semestre tiene cinco materias. Recolectamos las calificaciones de 401 estudiantes de pregrado que se inscribieron en el año académico 2014-15 y las usamos como conjunto de capacitación. Los datos contienen variables relacionadas con las calificaciones de los exámenes universitarios de los estudiantes en diversas materias que se impartieron en el primer y segundo semestre. La descripción del conjunto de datos utilizado se muestra en la Tabla 1. Para el conjunto de pruebas se utilizaron las notas obtenidas por los estudiantes en el tercer semestre.

La Tabla 2 muestra la descripción de los atributos del conjunto de datos utilizado.

La Figura 2 muestra el modelo de árbol generado por el conjunto de datos de entrenamiento usando REPTree.

El rendimiento de los siguientes clasificadores Meta, Bagging, Boosting, Dagging y Grading se examinó utilizando el entorno de flujo de conocimiento de Weka.

La Figura 3 muestra el árbol de muestra generado durante una iteración para el clasificador REPTree de Bagging. El atributo S3 se identifica como el nodo raíz.

Tabla 1: Descripción del conjunto de datos

Propiedades	Juego de entrenamiento	Juego de prueba
Número de instancias	401	80
Número de atributos	12	12
Atributos de clase	1	1
Número de alumnos aprobados	292	55
Número de estudiantes reprobados	109	25
Dimensión	401x13	80x13

Tabla 2: Descripción de atributos

Atributo	Descripción	Type	Rango (Conjunto de entrenamiento)	Rango (Conjunto de prueba)
Resultado	Resultado final	binominal	Aprobado (292), Reprobado (109)	Aprobado (55), Reprobado (25)
Sexo	Sexo	binominal	H (185), F (216) [15.000; 22.000] [15.000; 99.000]	F (50), M (30) [15.000; 20.000] [15.000; 98.000]
Envejecer	Envejecer	integer		
S1	Sujeto 1 puntos	integer	99.000] [12.000; 99.000]	[28.000; 97.000]
S2	Materia 2 puntos	integer	99.000] [14.000; 99.000]	[27.000; 95.000]
S3	Materia 3 puntos	integer	99.000] [18.000; 99.000]	[38.000; 99.000]
S4	Materia 4 puntos	integer	99.000] [11.000; 99.000]	[18.000; 99.000]
S5	Materia 5 puntos	integer	99.000] [11.000; 99.000]	[41.000; 99.000]
S6	Materia 6 puntos	integer	99.000] [15.000; 99.000]	[40.000; 97.000]
S7	Materia 7 puntos	integer	99.000] [12.000; 99.000]	[28.000; 96.000]
S8	Materia 8 puntos	integer	99.000] [15.000; 99.000]	[26.000; 94.000]
S9	Materia 9 puntos	integer		[28.000; 96.000]
S10	Materia 10 puntos	integer		

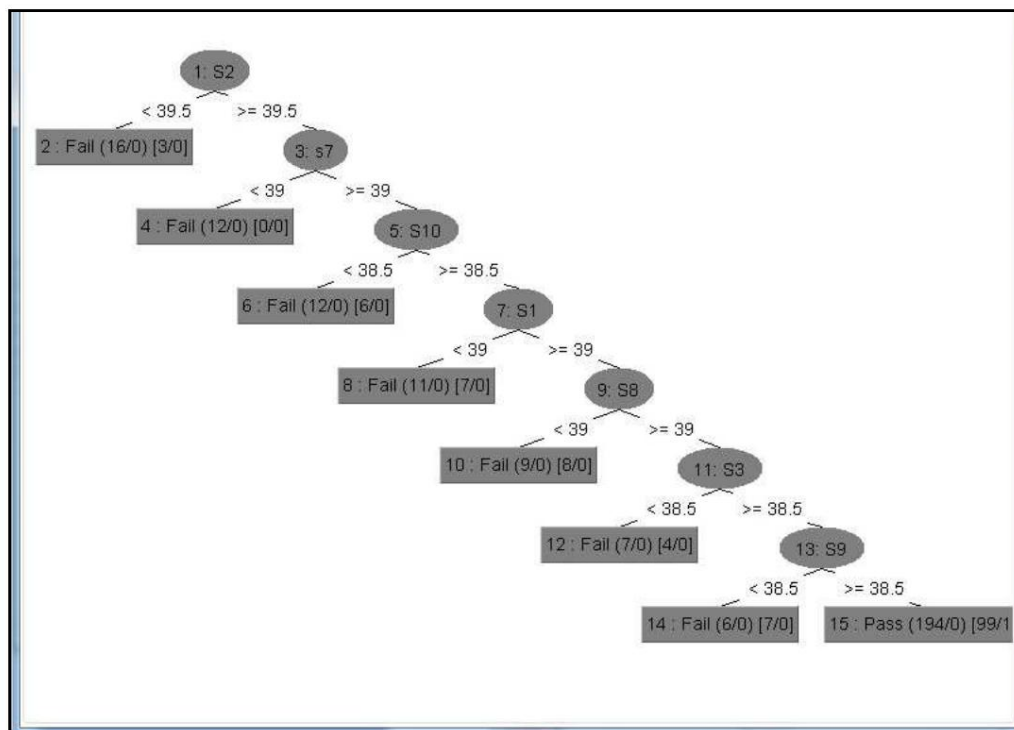


Fig. 2: Modelo generado por REPTree

```

S3 < 38: Fail (15/0) [8/0]
S3 >= 38
| s7 < 37.5: Fail (14/0) [4/0]
| s7 >= 37.5
| | S2 < 39: Fail (12/0) [2/0]
| | S2 >= 39
| | | S10 < 38: Fail (12/0) [5/0]
| | | S10 >= 38
| | | | S1 < 38.5: Fail (8/0) [6/0]
| | | | S1 >= 38.5
| | | | | S8 < 38: Fail (5/0) [4/0]
| | | | | S8 >= 38
| | | | | | S9 < 32.5: Fail (4/0) [3/0]
| | | | | | S9 >= 32.5: Pass (197/0) [102/3]

```

Size of the tree : 15

Fig. 3: Modelo generado por Bagging REPTree

```

s7 < 81.5
| S5 < 46.5
| | S5 < 40.5: Fail (7.64/0) [0.26/0]
| | S5 >= 40.5: Pass (27.32/0.52) [13.95/0.39]
| S5 >= 46.5
| | age < 16.5: Fail (165.76/6.59) [30.75/4.26]
| | age >= 16.5
| | | s7 < 39: Fail (12.79/0) [0.13/0]
| | | s7 >= 39: Pass (12.14/2.07) [17.83/1.03]
s7 >= 81.5
| S2 < 40.5: Fail (6.86/0) [0.13/0]
| S2 >= 40.5: Pass (95.34/1.16) [10.09/1.03]

```

Size of the tree: 13

Weight: 3.11

Fig. 4: Modelo generado por Boosting REPTree

```

S10 < 38.5: Fail (3/0) [1/0]
S10 >= 38.5: Pass (23/4) [13/3]

```

Size of the tree : 3

Fig. 5: Modelo generado por Dagging REPTree

```

S3 < 38.5: Fail (14/0) [2/0]
S3 >= 38.5
| S10 < 38.5: Fail (13/0) [5/0]
| S10 >= 38.5
| | S8 < 38.5: Fail (11/0) [7/0]
| | S8 >= 38.5
| | | S2 < 40.5
| | | S3 < 70.5: Fail (11/0) [8/2]
| | | S3 >= 70.5
| | | | S2 < 38.5: Fail (2/0) [2/0]
| | | | S2 >= 38.5: Pass (2/0) [1/0]
| | | S2 >= 40.5
| | | S9 < 38.5: Fail (8/0) [4/0]
| | | S9 >= 38.5
| | | | S1 < 38: Fail (7/0) [6/0]
| | | | S1 >= 38
| | | | | s7 < 37.5: Fail (6/0) [3/0]
| | | | | s7 >= 37.5: Pass (193/1) [96/1]
Size of the tree: 19

```

Fig. 6: Modelo generado por Grading REPTree

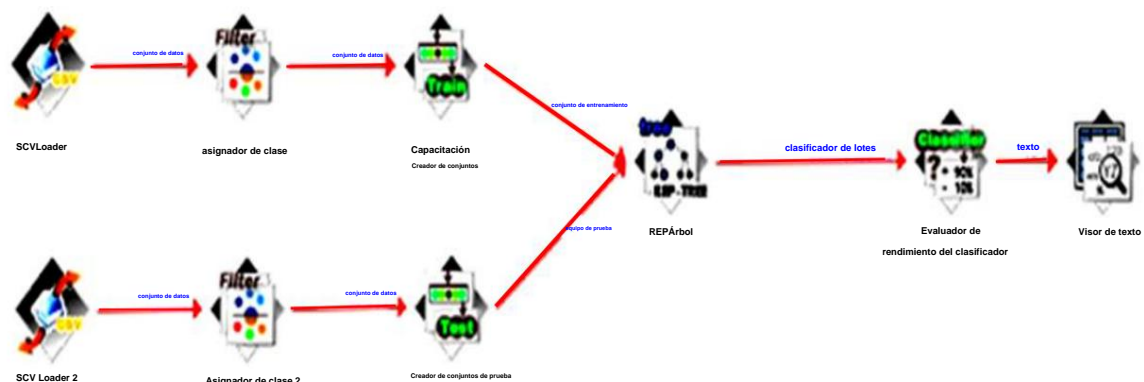


Fig. 7: Diseño del flujo de conocimiento de REPTree

La Figura 4 muestra un árbol modelo generado para el clasificador Boosting REPTree.

La Figura 5 muestra el árbol modelo obtenido para Dagging REPÁrbol.

La Figura 6 muestra el modelo generado por el metaclasificador de calificación.

Como se mencionó anteriormente, el clasificador del árbol de decisión Meta que utiliza el método REPTree se desarrolló utilizando el entorno de flujo de conocimiento de Weka. La configuración de parámetros disponible por defecto en la herramienta se utiliza para todos los algoritmos. En la Fig. 7 se muestra un diseño de flujo de conocimiento de muestra para el clasificador REP-Tree.

Las prestaciones de los clasificadores dependen de las características de los datos a clasificar. Los resultados de rendimiento de los algoritmos elegidos se utilizan para medir la exactitud, la precisión, la recuperación y la puntuación F. Las medidas de rendimiento obtenidas para el clasificador REPTree individual utilizado para el conjunto de entrenamiento y el conjunto de prueba se muestran en la Tabla 3.

El rendimiento de los clasificadores de árboles de metadecisiones se mide en términos de precisión, recuperación y puntuación F. Para evaluar el modelo de clasificación se han utilizado el conjunto de entrenamiento y el conjunto de prueba.

En la Tabla 3 se muestran los resultados de los algoritmos de árboles de metadecisiones diferentes: embolsado, impulso, d

se han utilizado para la comparación. Los resultados de la comparación se dan en la Tabla 4 en términos de precisión. La precisión más alta del 99,0 % se logró mediante el algoritmo de embolsado y potenciado de REPTree para el conjunto de datos de entrenamiento. Para el conjunto de prueba utilizado, la precisión más alta se obtiene para Boosting en lugar de Bagging.

Los valores de recuperación de los clasificadores de árbol de decisión meta empleados se muestran en la Tabla 5. Los valores de recuperación obtenidos prueban que los clasificadores de REPTree de Bagging y Boosting funcionaron mejor que los clasificadores de Dagging y Grading de REPTree tanto para el conjunto de entrenamiento como para el conjunto de prueba. La recuperación más alta del 99,0% se obtiene para los clasificadores REPTree de embolsado y potenciado. Para el conjunto de prueba, el mayor recuerdo se obtiene para el clasificador Boosting REPTree (97,5%).

Tabla 3: Desempeño del clasificador REPTree

Medidas	Conjunto de	Equipo de prueba
Precisión	entrenamiento 97,3	89,4
Recuerdo	97,3 97,2	87,5
Puntuación F		86,3

Tabla 4: Precisión de los clasificadores del árbol de metadecisiones

Metaclassificadores (REPTree)	Conjunto de	Equipo de prueba
Harpillera	entrenamiento 99.0	96.4
impulsar	99.0	97.6
Dagging	80.3	50.8
calificación	98.0	92.5

Tabla 5: Recuperación de clasificadores de árbol de decisión Meta

Metaclassificadores (REPTree)	Conjunto de	Equipo de prueba
Harpillera	entrenamiento 99,0	96,3
impulsar	99,0 73,1	97,5
Dagging		71,3
calificación	98.0	92.5

Tabla 6: Puntuación F de los clasificadores del árbol de decisión Meta

Metaclassificadores (REPTree)	Conjunto de	Equipo de prueba
Harpillera	entrenamiento 99,0	96,2
impulsar	99,0	97,5
Dagging	61,9	51,4
calificación	98.0	92.5

Tabla 7: Comparación con trabajos anteriores

Autor	Método de predicción	Precisión obtenida
Wolff et al. (2014)	Red bayesiana	89,6% 92,3% 91,9%
Romero et al. (2013)	Máquinas de vectores soporte	93,6% 97,5% (embolsado)
Ramesh et al. (2013)	Árbol de decisión	
Zhang et al. (2015)	Redes neuronales	
Nuestro enfoque	método de conjunto	

Test output				
Tester: weka.experiment.PairedTTester				
Analysing: Percent_correct				
Datasets: 1				
Resultsets: 4				
Confidence: 0.05 (two tailed)				
Sorted by: -				
Date: 8/14/16 2:40 PM				
Dataset	(1) meta.Bagging ' (2) meta.AdaBoo (3) meta.Daggin (4) meta.Gradin			
student-train-full	(100)	98.41 (2.17)	98.35 (2.05)	73.27 (1.45) * 96.22 (4.50) *
Average		98.41	98.35	73.27 96.22
		(v/ /*)	(0/1/0)	(0/0/1) (0/0/1)
Key:				
(1) meta.Bagging '-P 100 -S 1 -num-slots 1 -I 10 -W trees.REPTree -- -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0' -115879962237199703				
(2) meta.AdaBoostM1 '-P 100 -S 1 -I 10 -W trees.REPTree -- -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0' -1178107808933117974				
(3) meta.Dagging '-F 10 -S 1 -W trees.REPTree -- -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0' 4560165876570074309				
(4) meta.Grading '-X 10 -M \"rules.ZeroR \" -S 1 -num-slots 1 -B \"trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0\" 52078379478900811'				

Fig. 8: Comparación de precisiones usando Weka Experimenter

Los resultados de F-Score para los clasificadores del árbol de metadecisiones empleados se muestran en la Tabla 6. El F-score es una combinación ponderada de Bagging y Boosting. Por lo tanto, se observan observaciones casi similares de precisión y recuperación para la puntuación F de los clasificadores del árbol de metadecisiones. Los valores de puntuación F también muestran que los clasificadores Bagging y Boosting de REPTree funcionan mejor que Dagging and Grading of REPTree tanto para el conjunto de entrenamiento como para el conjunto de prueba.

La precisión de los clasificadores del árbol de decisión Meta se calcula y compara utilizando el experimentador Weka. El resultado de precisión obtenido muestra que el embolsado se muestra mejor precisión que los otros clasificadores de árbol de meta decisión (Fig. 8). El clasificador REPTree impulsor ocupa el siguiente lugar en la jerarquía. También se observan resultados similares para el conjunto de prueba.

Los resultados obtenidos en este trabajo de investigación se comparan con el trabajo existente y los resultados se tabulan en la Tabla 7. Los resultados muestran que el método conjunto demuestra ser mejor que otros métodos individuales empleados anteriormente para predecir los resultados de los estudiantes en las instituciones educativas.

Conclusión

Predecir el desempeño del estudiante es la forma más efectiva para que los dedicados y los alumnos mejoren sus procesos de enseñanza y aprendizaje. Se pueden sacar mejores inferencias con el enfoque de clasificación, lo que da como resultado una mejor predicción de si un estudiante aprobará o reprobará en un curso. Es necesario realizar más análisis para comprender mejor y mejorar estos resultados. Además de problemas estudiados en este trabajo, sería interesante predecir un intervalo para un grado. Este método ayudará a las instituciones educativas a monitorear el desempeño de los estudiantes de manera efectiva y sistemática. Por último, este modelo puede ayudar a los educadores a comprender a los alumnos, identificar a los alumnos débiles, mejorar los procesos de aprendizaje y reducir las tasas de fracaso académico. También puede ayudar a los administradores a mejorar los resultados del sistema de aprendizaje.

En nuestro trabajo futuro, nos centraremos más en analizar las características de comportamiento en el modelo de desempeño de los estudiantes. Este resultado puede resultar ser un modelo predictivo más realista. Se podría realizar alguna optimización utilizando un método de selección de parámetros, como la selección de características. En conclusión, el metaanálisis sobre la predicción del rendimiento de un estudiante nos ha inspirado para realizar más investigaciones para ser aplicadas en diversas instituciones educativas.

Contribuciones del autor

Shanthini, A.: Participó en todos los datos del experimento y parte de la redacción.

G. Vinodhini: Análisis e investigación de datos.

RM Chandrasekaran: Plan de investigación y redacción.

Ética

Este artículo es original y contiene material inédito. El autor correspondiente confirma que todos los demás autores han leído y aprobado el manuscrito y que no hay problemas éticos involucrados.

Referencias

- Amrieh, EA, T. Hamtini e I. Aljarah, 2016. Extracción de datos educativos para predecir el rendimiento académico de los estudiantes mediante métodos de conjunto. En t. j Aplicación de la teoría de bases de datos, 9: 119-136. DOI: 10.14257/ijdata.2016.9.8.13 Arsad, PM, N. Bunyamin y JLA Manan, 2013. Modelo de predicción del rendimiento de los estudiantes de una red neuronal (NNSPPM). Actas de la Conferencia internacional IEEE sobre instrumentación, medición y aplicaciones inteligentes, del 25 al 27 de noviembre, IEEE Xplore Press, Kuala Lumpur, Malasia, págs. 1 a 5. DOI: 10.1109 / ICSIMA.2013.6717966 Asogbon, MG, OW Samuel, MO Omisore y BA Ojokoh, 2016. Un enfoque de máquina de vectores de soporte multiclasa para la predicción del rendimiento académico de los estudiantes. En t. J. Curr Multidisciplinar. Res., 4: 210-215. Buldu, A. y K. Ucgun, 2010. Aplicación de minería de datos sobre datos de estudiantes. proc. Soc. Comportamiento Sci., 2: 5251-5259. DOI: 10.1016/j.sbspro.2010.03.855 Chen, SM y SM Bai, 2010. Uso de técnicas de minería de datos para construir automáticamente mapas conceptuales para sistemas de aprendizaje adaptativo. Sistema experto Aplicación, 37: 4496-4503. DOI: 10.1016/j.eswa.2009.12.060 Delen, D., 2010. Un análisis comparativo de las técnicas de aprendizaje automático para la gestión de la retención de estudiantes. Decir Sistema de soporte, 49: 498-506. DOI: 10.1016/j.dss.2010.06.003 Hien, NTN y P. Haddawy, 2007. Un sistema de apoyo a la decisión para evaluar solicitudes de estudiantes internacionales. Actas de la 37.ª Conferencia Anual de Fronteras en Educación-Ingeniería Global: Conocimiento, 2007. IEEE Xplore Press, Milwaukee, WI, EE. UU., pp: F2A-1-F2A-6. DOI: 10.1109/FIE.2007.4417958 Illanas Vila, AI, JR Calvo Ferrer, FJ Gallego-Durán y LF Largo, 2013. Predicción del rendimiento de los estudiantes en la traducción de lenguas extranjeras con un juego serio. Actas de la Conferencia INTED, 4-6 de marzo, Valencia, España, pp: 52-59. Marquez Vera, C., A. Cano, C. Romero, AYM Noaman and H. Mousa Fardoun et al., 2016. Predicción temprana de la deserción escolar usando minería de datos: un estudio de caso con estudiantes de secundaria. Sistema experto, 33: 107-124. DOI: 10.1111 / exsy.12135

- Nespereira, CG, E. Elhariri, N. El-Bendary, AF Vilas y RPD Redondo, 2015. Enfoque de clasificación basado en el aprendizaje automático para predecir el rendimiento de los estudiantes en el aprendizaje combinado. Actas de la 1ra Conferencia Internacional sobre Informática y Sistemas Inteligentes Avanzados, 28-30 de noviembre, Springer, Beni Suef, Egipto, pp: 47-56. DOI: 10.1007/978-3-319-26690-9_5
- Petkovic, D., K. Okada, M. Sosnick, A. Iyer y S. Zhu et al., 2012. Trabajo en curso: un enfoque de aprendizaje automático para la evaluación y predicción de la eficacia del trabajo en equipo en la programación de software. Actas de la Conferencia Frontiers in Education, del 3 al 6 de octubre, IEEE Xplore Press, Seattle, WA, EE. UU., págs. 1-3. DOI: 10.1109 / FIE.2012.6462205
- Ramesh, V., P. Parkavi y K. Ramar, 2013. Predecir el desempeño de los estudiantes: un enfoque estadístico y de minería de datos. En t. J. Cómputo. Aplíc., 63: 35-39. DOI: 10.5120 / 10489-5242
- Romero, C., MI López, JM Luna and S. Ventura, 2013. Predicción del desempeño final de los estudiantes a partir de la participación en foros de discusión en línea. compeduc., 68: 458-472. DOI: 10.1016/j.compedu.2013.06.009
- Schalk, PD, DP Wick, PR Turner y MW Ramsdell, 2011. Evaluación predictiva del desempeño de los estudiantes para orientación estratégica temprana. Actas de la Conferencia Frontiers in Education, del 12 al 15 de octubre, IEEE Xplore Press, Rapid City, SD, EE. UU., pp: S2H-1-S2H-5. DOI: 10.1109 / FIE.2011.6143086
- Sorour, SE, K. Goda y T. Mine, 2015. Estimación del desempeño de los estudiantes basada en modelos de temas considerando una variedad de lecciones. Actas de la Conferencia Internacional sobre Inteligencia Artificial en la Educación, 22-26 de junio, Springer, pp: 790-793. DOI: 10.1007/978-3-319-19773-9_117
- Strecht, P., L. Cruz, C. Soares, J. Mendes-Moreira and R. Abreu, 2015. Estudio comparativo de algoritmos de clasificación y regresión para modelar el rendimiento académico de los estudiantes. Actas de la 8ª Conferencia Internacional sobre Minería de Datos Educativos, del 26 al 29 de junio, en Madrid, España.
- Wolff, A., Z. Zdrahal, D. Herrmannova y P. Knoth, 2014. Predicción del rendimiento de los estudiantes a partir de fuentes de datos combinadas. En: Minería de datos educativos, Peña-Ayala, A. (Ed.), Springer International Publishing, pp: 175-202.
- Yoo, J. y J. Kim, 2014. ¿Puede la participación en discusiones en línea predecir el desempeño del proyecto grupal? Investigar los roles de las características lingüísticas y los patrones de participación. En t. J. Inteligencia Artificial. Educ., 24: 8-32. DOI: 10.1007/s40593-013-0010-8
- Zhang, W., S. Zhang y S. Zhang, 2015. Predicción de la calificación de la tesis de graduación mediante SVM. En t. J. Intel. Informar. Proceso., 5: 60-68.