

Listas de contenidos disponibles en [ScienceDirect](#)

Computadoras en el comportamiento humano

página de inicio de la revista: www.elsevier.com/locate/comphumbeh

Artículo de longitud completa

Evaluación de la efectividad de las técnicas educativas de minería de datos para la predicción temprana del fracaso académico de los estudiantes en cursos de introducción a la programación



Evandro B. Costa ^a, Baldoino Fonseca ^b, ^a, marcelo almeida santana ^a,
 Ferreira de Araújo ^c, Joilson Rego ^d, ^c Fabrisia ^d

^a Universidad Federal de Alagoas (UFAL), Brasil^b Instituto Federal de Alagoas (IFAL), Brasil^c Universidad Federal de Campina Grande, Brasil^d Universidad Federal de Rio Grande do Norte (UFRN), Brasil

información del artículo

Historial del artículo:

Recibido el 13 de enero de 2016

Recibido en forma revisada

16 enero 2017

Aceptado el 26 de enero de 2017

Disponible en línea el 4 de febrero de 2017

Palabras clave:

inteligencia artificial en la educación

planificador instructivo automático

Predicción automática

Minería de datos educativos

Entorno de aprendizaje interactivo

Modelado de alumnos

abstracto

Los datos sobre las altas tasas de fracaso de los estudiantes en los cursos de introducción a la programación han alarmado a muchos educadores y plantean una serie de preguntas importantes con respecto a los aspectos de predicción. En este artículo, presentamos un estudio comparativo sobre la efectividad de las técnicas educativas de minería de datos para predecir tempranamente que los estudiantes probablemente reprobarán cursos de introducción a la programación. Aunque varios trabajos han analizado estas técnicas para identificar los fracasos académicos de los estudiantes, nuestro estudio difiere de los existentes en lo siguiente: (i) investigamos la eficacia de tales técnicas para identificar a los estudiantes con probabilidades de fracasar en una etapa lo suficientemente temprana como para tomar medidas para reducir la tasa de fracaso; (ii) analizamos el impacto de las tareas de preprocesamiento de datos y ajuste de algoritmos, sobre la efectividad de las técnicas mencionadas. En nuestro estudio, evaluamos la efectividad de cuatro técnicas de predicción en dos fuentes de datos diferentes e independientes sobre cursos de introducción a la programación disponibles en una universidad pública brasileña: una proviene de la educación a distancia y la otra en el campus. Los resultados mostraron que las técnicas analizadas en nuestro estudio son capaces de identificar tempranamente a los estudiantes con probabilidad de reprobar, la efectividad de algunas de estas técnicas mejora después de aplicar el preprocesamiento de datos y/o el ajuste fino de los algoritmos, y la técnica de la máquina de vectores de soporte supera a la otros de forma estadísticamente significativa.

© 2017 Elsevier Ltd. Todos los derechos reservados.

1. Introducción

Los índices alarmantes de fracasos académicos de los estudiantes, a lo largo de los años, en los cursos introductorios de programación de las universidades (Benndsen & Caspersen, 2007; Watson & Li, 2014) han preocupado a los educadores. Los estudios (Hanks et al., 2004; Iepsen et al., 2013; Tan, Ting, & Ling, 2009) muestran que los estudiantes enfrentan muchas dificultades durante sus actividades de programación de tal manera que muchos de ellos terminan suspendiendo o abandonando el programa. curso en alguna etapa inicial.

En el contexto anterior, un problema relevante es la capacidad de predecir con precisión la probabilidad de que los estudiantes suspendan en los cursos introductorios de programación en una etapa lo suficientemente temprana como para posibilitar

intervenciones pedagógicas a tomar para evitar los fracasos de los estudiantes. Para abordar este problema, algunos trabajos (Arora, Singhal, & Bansal, 2014; Bayer et al., 2012; Manhães et al., 2014; Marquez Vera, Morales, & Soto, 2013; Martinho, Nunes, & Minussi, 2013; Sim et al., 2006; Watson, Li, & Godwin, 2013) han propuesto y analizado el uso de técnicas de Minería de Datos Educativos (EDM) para predecir el fracaso académico de los estudiantes. Sin embargo, en general, estos trabajos no se ocupan de dos preguntas importantes: (i) ¿qué tan efectivas son las técnicas de EDM para identificar tempranamente a los estudiantes con probabilidades de reprobar?; y (ii) hacer el preprocesamiento de datos (Hu, 2003; Crone et al., 2006; Zaki & Jr.WM, 2014) y el ajuste de algoritmos (Gunawan et al., 2011; Hutter, Hoos, Leyton-Brown, & Stützle, 2009) impactan la efectividad de las técnicas de EDM?

Con el fin de responder a las preguntas mencionadas anteriormente, presentamos un estudio comparativo sobre la eficacia de las técnicas de EDM para predecir de forma temprana que los estudiantes probablemente reprobarán cursos de introducción a la programación. Dada la cantidad de técnicas de EDM existentes disponibles (Caruana &

* Autor correspondiente.

Correos electrónicos: evandro@ic.ufal.br (EB Costa), baldoino@ic.ufal.br (B. Fonseca), marceloalmeidasantana@gmail.com (MA Santana), fabrisia.araujo@gmail.com (FF de Araújo), joilreg@gmail.com (J.Rego).

Niculescu-Mizil, 2006), utilizamos los siguientes clasificadores: Neural Networks (Nürnberger et al., 2002; Rumelhart, Hinton, & Williams, 1988), Decision Tree (Breiman et al., 1984; Salzberg, 1994), Support Vector Machine (SVM) (Cortes & Vapnik, 1995; Vapnik, 1995) y Naive Bayes (Domingos & Pazzani, 1997). Estas técnicas han sido ampliamente investigadas por trabajos de EDM existentes (Wu et al., 2008) y han presentado resultados interesantes.

En nuestro estudio, utilizamos la medida *f* (Han et al., 2011) para evaluar la eficacia de las técnicas seleccionadas en dos fuentes de datos diferentes e independientes sobre dos cursos de introducción a la programación disponibles en una universidad pública brasileña: uno proviene de la educación a distancia, y el otro del campus. El experimento se realizó considerando el preprocesamiento de estas fuentes de datos y la puesta a punto de las técnicas analizadas.

Los resultados mostraron que las técnicas analizadas en nuestro estudio son capaces de identificar tempranamente a los estudiantes con probabilidad de reprobación, y demostraron que las tareas de preprocesamiento de datos y ajuste de algoritmos influyen en la efectividad de estas técnicas. La técnica SVM superó a las demás al predecir con un 92% y un 83% de efectividad los fracasos de los estudiantes que han realizado al menos el 50% de los cursos a distancia o presencial, respectivamente.

Este artículo está organizado de la siguiente forma: En la sección 2 presentamos el método aplicado en nuestro experimento. En la Sección 3 presentamos los resultados y discusiones del experimento. En la Sección 4 discutimos algunos trabajos similares. Las conclusiones y el trabajo futuro se presentan en la Sección 5.

2. Método

El objetivo general de este estudio es comparar la efectividad de las técnicas EDM existentes para la identificación temprana de estudiantes con probabilidad de reprobación con alta precisión. Esta sección está organizada de la siguiente manera. La Sección 2.1 plantea cuatro preguntas de investigación que impulsan nuestra evaluación. La Sección 2.2 presenta las fuentes de datos y las técnicas de EDM que hemos analizado en nuestro experimento. Las secciones 2.3 y 2.4 indican las herramientas y métricas, respectivamente, que hemos utilizado al realizar el experimento y, finalmente, la sección 2.5 presenta algunos detalles sobre los pasos y las configuraciones utilizadas para realizar el experimento.

2.1. Planificación

Nuestro estudio comparativo está guiado por las siguientes preguntas de investigación:

Pregunta 1. ¿Qué tan efectivas son las técnicas de EDM para identificar tempranamente a los estudiantes con probabilidades de reprobación?

Nuestro objetivo con la Pregunta 1 es evaluar la efectividad de las técnicas de EDM que han sido utilizadas por los enfoques existentes para identificar tempranamente a los estudiantes con probabilidades de reprobación. Para responder a la Pregunta 1, realizamos estas técnicas en dos fuentes de datos diferentes y luego usamos la medida *F* para evaluar la efectividad de tales técnicas

Pregunta 2. ¿Es el preprocesamiento de datos capaz de aumentar la eficacia de las técnicas de EDM?

La Pregunta 2 tiene como objetivo analizar si la efectividad de las técnicas de EDM aumenta después de realizar el preprocesamiento de datos. Para responder a la Pregunta 2, realizamos un procesamiento previo de las dos fuentes de datos utilizadas en este experimento, luego aplicamos las técnicas de EDM en estas fuentes de datos. Posteriormente, evaluamos la efectividad de estas técnicas y comparamos dichos resultados con la efectividad obtenida al realizar las mismas técnicas.

en los datos sin el preprocesamiento.

Pregunta 3. ¿Puede el ajuste fino de los algoritmos aumentar aún más la eficacia de las técnicas de EDM?

La Pregunta 3 tiene como objetivo analizar si la efectividad de las técnicas de EDM aumenta aún más después de realizar el ajuste fino de sus parámetros. Para responder a la Pregunta 3, realizamos un ajuste fino de las técnicas de EDM, luego realizamos el ajuste fino de las técnicas en la fuente de datos preprocesada, así como también evaluamos la efectividad de las técnicas y comparamos su efectividad con los resultados. obtenido mediante la realización de las técnicas de electroerosión sin el ajuste fino.

Pregunta 4. Después de realizar el preprocesamiento de datos y el ajuste fino de los algoritmos, ¿cuáles de las técnicas de EDM son más efectivas para la identificación temprana de estudiantes con probabilidades de reprobación?

La Pregunta 4 tiene como objetivo encontrar las técnicas más efectivas para la identificación temprana de los estudiantes con probabilidades de reprobación. Para responder a la Pregunta 4, analizamos la efectividad de las técnicas de EDM después de realizar el ajuste fino de sus parámetros y el procesamiento previo de las fuentes de datos.

2.2. Selección de fuentes de datos y técnicas de EDM

En este experimento hemos analizado dos fuentes de datos extraídas de cursos de introducción a la programación realizados en el campus o la educación a distancia. A continuación se ofrece una breve descripción de estas dos fuentes de datos:

(Educación a Distancia) La primera fuente de datos contiene información de 262 estudiantes de pregrado que tomaron el curso de introducción a la programación realizado en una modalidad de educación a distancia en nuestra universidad en el año 2013 durante 10 semanas. En este curso los alumnos fueron evaluados semanalmente de acuerdo a sus actividades más dos exámenes que se aplicaron en la quinta y última semana del curso. Estas actividades y exámenes se aplicaron a través de un sistema en línea utilizado en nuestra universidad.

Esta fuente de datos contiene la siguiente información sobre los estudiantes: edad, sexo, estado civil, ciudad, ingresos, matrícula de estudiantes, período, clase, semestre, campus, frecuencia de acceso de los estudiantes al sistema, participación en el foro de discusión, cantidad de archivos recibidos y visualizados, uso de las herramientas educativas que brinda el sistema como blog, glosario, quiz, wiki, mensaje, año de ingreso al curso, estado de la disciplina y desempeño de los estudiantes en las actividades semanales y exámenes.

(Presencial) La segunda fuente de datos contiene información sobre 161 estudiantes que tomaron el curso de introducción a la programación realizado en el campus de nuestra universidad en 2014, durante 16 semanas. En este curso los estudiantes fueron evaluados semanalmente de acuerdo a sus actividades más cuatro exámenes que se aplicaron en la cuarta, octava, duodécima y decimosexta semana del curso.

La fuente de datos contiene la siguiente información de los estudiantes: edad, sexo, estado civil, ciudad, ingresos, registro del estudiante, período, clase, semestre, campus, año de inscripción en el curso, estado en la disciplina, cantidad de ejercicio realizado por el estudiante, número de ejercicios correctos, y desempeño de los estudiantes en las actividades semanales y exámenes.

Nuestro objetivo principal es evaluar la efectividad de las técnicas de EDM para predecir la probabilidad de que los estudiantes fallen en una etapa lo suficientemente temprana como para apoyar futuras intervenciones pedagógicas que se deben tomar para evitar

Descargar versión en inglés:

<https://daneshyari.com/en/article/4937176>

Descargar versión persa:

<https://daneshyari.com/article/4937176>

[Daneshyari.com](https://daneshyari.com)