



Memòries del Programa de XARXES-I³CE de qualitat,
innovació i investigació en docència universitària.
Convocatòria 2018-19

Memorias del Programa de REDES-I³CE de calidad,
innovación e investigación en docencia universitaria.
Convocatoria 2018-19

Rosabel Roig-Vila (Coord.)
Jordi M. Antolí Martínez, Asunción Lledó
Carreres, Neus Pellín Buades (Eds.)



Memòries del Programa de Xarxes-I3CE
de qualitat, innovació i investigació en
docència universitària.
Convocatòria 2018-19

*Memorias del Programa de Redes-I3CE
de calidad, innovación e investigación
en docencia universitaria.
Convocatoria 2018-19*

Rosabel Roig-Vila (Coord.), Jordi M. Antolí Martínez, Asunción
Lledó Carreres, Neus Pellín Buades (Eds.)

Memòries de les xarxes d'investigació en docència universitària pertanyent al Programa Xarxes-I3CE d'Investigació en docència universitària del curs 2018-19 / *Memorias de las redes de investigación en docencia universitatira que pertenece al Programa Redes -I3CE de investigación en docencia universitaria del curso 2018-19*

Organització: Institut de Ciències de l'Educació (Vicerectorat de Qualitat i Innovació Educativa) de la Universitat d'Alacant/ *Organización: Instituto de Ciencias de la Educación (Vicerrectorado de Calidad e Innovación Educativa) de la Universidad de Alicante*

Edició / *Edición*: Rosabel Roig-Vila (Coord.), Jordi M. Antolí Martínez, Asunción Lledó Carreres, Neus Pellín Buades (Eds.)

Comité tècnic / *Comité técnico*: Neus Pellín Buades

Revisió i maquetació: ICE de la Universitat d'Alacant/ *Revisión y maquetación*: ICE de la Universidad de Alicante

Primera edició: / *Primera edición*: Novembre 2019

© De l'edició/ *De la edición*: Rosabel Roig-Vila , Jordi M. Antolí Martínez, Asunción Lledó Carreres & Neus Pellín Buades.

© Del text: les autores i autors / *Del texto: las autoras y autores*

© D'aquesta edició: Institut de Ciències de l'Educació (ICE) de la Universitat d'Alacant / *De esta edición: Instituto de Ciencias de la Educación (ICE) de la Universidad de Alicante*

ice@ua.es

ISBN: 978-84-09-15746-4

Qualsevol forma de reproducció, distribució, comunicació pública o transformació d'aquesta obra només pot ser realitzada amb l'autorització dels seus titulars, llevat de les excepcions previstes per la llei. Adreceu-vos a CEDRO (Centro Español de Derechos Reprográficos, www.cedro.org) si necessiteu fotocopiar o escanejar algun fragment d'aquesta obra. / *Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra sólo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la ley. Diríjase a CEDRO (Centro Español de Derechos Reprográficos, www.cedro.org) si necesita fotocopiar o escanear algún fragmento de esta obra.*

Producció: Institut de Ciències de l'Educació (ICE) de la Universitat d'Alacant / *Producción: Instituto de Ciencias de la Educación (ICE) de la Universidad de Alicante*

EDITORIAL: Les opinions i continguts dels resums publicats en aquesta obra són de responsabilitat exclusiva dels autors. / *Las opiniones y contenidos de los resúmenes publicados en esta obra son de responsabilidad exclusiva de los autores.*

48. Desarrollo de un servicio online para el uso de técnicas de aprendizaje automático orientadas a la detección de anomalías en la evaluación docente

Antonio-Javier Gallego, Juan Ramón Rico-Juan, Jorge Calvo-Zaragoza,
Francisco José Castellanos Regalado, David Rizo Valero

*jgallego@dlsi.ua.es, juanramonrico@ua.es, jcalvo@dlsi.ua.es,
fcastellanos@dlsi.ua.es, drizo@dlsi.ua.es*

*Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante,
Carretera San Vicente del Raspeig s/n, Alicante, 03690, Spain*

RESUMEN (ABSTRACT)

Uno de procesos más importantes en casi todos los modelos de enseñanza universitaria es la evaluación. Los criterios que se establecen en una asignatura orientan la forma en la que se obtiene la calificación final del alumno. Por este motivo es importante realizar un seguimiento continuado del aprendizaje del estudiante y de sus calificaciones, permitiendo de este modo la detección de anomalías para proceder con una intervención inmediata que permita corregir la situación. Normalmente, en los primeros cursos universitarios el número de alumnos es elevado, lo que redundaría en el detrimento del seguimiento que se le puede realizar a los estudiantes por parte del profesor. En el trabajo realizado en esta red docente se propone un sistema para predecir la calificación de un estudiante en una determinada actividad, de forma que se notifique al profesor cuando la calificación se aleje del valor predicho. Para esto se ha realizado un estudio de 24 algoritmos de inteligencia artificial, seleccionando finalmente los más adecuados para el caso de estudio realizado. Los resultados experimentales muestran la utilidad del método propuesto y cómo los algoritmos basados en máquinas de vectores soporte o los de aumentado de gradiente extremo son los que mejores resultados obtienen.

Palabras clave:

Aprendizaje automático, detección de anomalías, predicción de notas, evaluación.

1. INTRODUCCIÓN

Cada vez es más habitual encontrar publicaciones que exploran la posibilidad de aplicar técnicas de aprendizaje automático (área de la inteligencia artificial que estudia cómo pueden aprender los ordenadores a partir de datos) para prever problemas e intentar corregirlos antes de que sucedan (Barnes, 2017). Por ejemplo, predecir el fracaso académico de los estudiantes en los cursos de programación introductoria (Costa, 2017) o predecir si un estudiante finalizará satisfactoriamente o no su título universitario (Daud, 2017).

Si consideramos los estudios universitarios actuales no cabe duda que uno de los procesos más importantes en los modelos de enseñanza es la evaluación. Los criterios establecidos orientan la forma en la que los alumnos obtienen sus calificaciones parciales así como su nota final. A su vez, la evaluación continua supone un seguimiento del aprendizaje del estudiante que facilita la detección de anomalías en sus calificaciones en fases tempranas y permite la intervención inmediata para corregir la situación.

Normalmente, en los primeros cursos de estudios universitarios, el número de alumnos es elevado y ello redundaría en el detrimento del seguimiento directo que se puede realizar a los estudiantes por parte del profesor. Por lo tanto, un sistema para la detección temprana de anomalías en la evaluación continua basado en técnicas de aprendizaje automático tendría la ventaja de ayudar al profesor a identificar qué alumnos pueden tener dificultades con la asignatura. Básicamente, el sistema aprendería de las experiencias del profesor (histórico de calificaciones de cursos anteriores) para identificar posibles problemas futuros.

2. MÉTODO

El método propuesto se divide en dos fases (ver figura 1). En la primera fase se entrena el sistema predictivo utilizando el histórico de calificaciones de alumnos de cursos anteriores para la misma asignatura. La segunda fase se divide a su vez en dos tareas específicas: (1) En primer lugar se utiliza el sistema predictivo entrenado para obtener la nota esperada para los alumnos del curso actual; (2) A continuación se analizan las predicciones obtenidas para calcular la diferencia con las calificaciones reales de los alumnos.

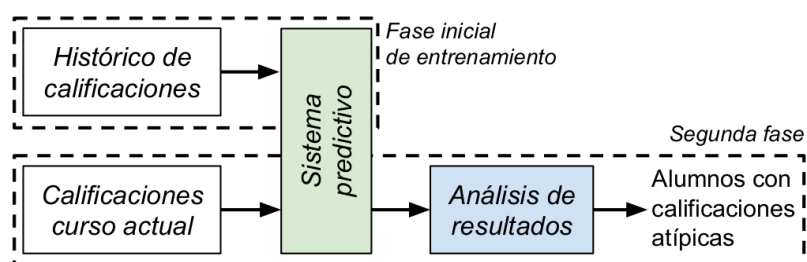


Figura 1: Esquema del método propuesto.

Con esta información se puede crear un sistema que notifique al profesor aquellos alumnos o alumnas cuyas calificaciones difieran de lo esperado un cierto umbral, seleccionando únicamente las que se consideren atípicas, es decir, aquellas que se encuentren por debajo de un percentil establecido (en nuestro sistema hemos utilizado el 10%).

Los sistemas predictivos basados en aprendizaje automático requieren que se defina un conjunto de características que representen el problema al que se pretende dar solución. En el caso de estudio propuesto, las características básicas a tener en cuenta para cada alumno son: las calificaciones previas obtenidas, el grupo al que pertenece (mañana o tarde) y su género.

En la literatura existen multitud de algoritmos predictivos, por lo que resulta de especial interés seleccionar aquellos que minimicen el error tanto como sea posible, permitiendo optimizar la precisión de las estimaciones. En aprendizaje automático, los algoritmos que predicen un valor numérico continuo son conocidos como modelos de regresión. Por ello hemos seleccionado una serie de algoritmos basados en diferentes estrategias para abarcar el mayor número de aproximaciones al problema y así poder evaluar su desempeño en el caso de estudio propuesto.

En primer lugar se ha utilizado la herramienta de análisis de datos WEKA (v3.7.12) para realizar los experimentos. En concreto se han considerado los siguientes métodos (el listado se indica utilizando la nomenclatura y clasificación de la librería): funciones (*GaussianProcesses*, *IsotonicRegression*, *LinearRegression*, *LeastMedSq*, *PaceRegression*, *RBFFNetwork*, *RBFRRegressor* y *SMOreg*); árboles (*RandomForest* y *M5P*); basado en prototipos (*lazy*, como *LWL*); y meta-algoritmos (*AdditiveRegression*, *RandomSubSpace* y *RandomCommittee*).

También se han usado paquetes del lenguaje Python como Sklearn (0.19), XGBoost (0.6), LightGBM (2.0.7) y Keras (2.0.8), para evaluar los siguientes algoritmos: Algoritmos

lineales (*LinearRegression*, *Ridge* y *BayesianRidge*); árboles (*DecisionTree*); regla de los vecinos más cercanos (*KNeighbors*); boosting (*XGBoost* *XGB/xgboost*, *LightGBM* *LGBM/lightgbm*); support Vector Machine (*SVR/SVM*); y Artificial Neural Networks (ANN) con Keras (*ANN-dense con 1 capa y 32 neuronas*, *ANN-avg con GlobalAveragePolling, 32 neuronas y 0.2 de dropout*).

3. RESULTADOS

Los datos a evaluar han sido obtenidos a partir de las calificaciones de cuatro prácticas (P1, P2, P3 y P4) de 751 alumnos correspondientes a cuatro cursos académicos completos (2013-2016) de una asignatura de informática introductoria. Para validar cada uno de los algoritmos de regresión seleccionados se ha utilizado la técnica de validación cruzada sobre 10 particiones (10-CV en adelante) aplicada habitualmente para este tipo de tareas. Para medir la calidad de los resultados obtenidos a partir de los algoritmos predictivos se ha utilizado el *error absoluto medio* (referenciado como MAE en adelante).

Según la calificación de la práctica que se quiera predecir (P1, P2, P3 y P4) partiremos de un conjunto de datos inicial distinto. De esta forma establecemos cuatro escenarios o grupos de experimentos, como se indica a continuación: (1) Grupo, Género \rightarrow P1, (2) Grupo, Género, P1 \rightarrow P2, (3) Grupo, Género, P1, P2 \rightarrow P3, y (4) Grupo, Género, P1, P2, P3 \rightarrow P4.

Tabla 1: Resultados de los errores absolutos medios (sobre 10 puntos) de los algoritmos evaluados con la técnica de validación cruzada. Sombreado en verde se muestra el primer cuartil y en rojo el último cuartil por columna. Un valor menor representa un mejor resultado.

Paquete	Algoritmo	P1	P2	P3	P4	Media
weka.functions	GaussianProcesses	1,22	1,17	0,85	0,42	0,92
	IsotonicRegression	1,21	1,2	0,87	0,56	0,96
	LinearRegression(wk)	1,21	1,17	0,85	0,42	0,91
	LeastMedSq	1,06	1,13	0,85	0,42	0,87
	PaceRegression	1,21	1,16	0,85	0,42	0,91
	RBFNetwork	1,23	1,2	0,88	0,64	0,99
	RBFRegressor	1,22	1,16	0,84	0,43	0,91
	SMOreg	1,06	1,1	0,83	0,42	0,85
weka.trees	RandomForest	1,22	1,16	0,85	0,41	0,91
	M5P	1,21	1,17	0,85	0,42	0,91
weka.lazy	LWL	1,21	1,18	0,86	0,54	0,95
weka.meta	AdditiveRegression	1,22	1,17	0,83	0,43	0,91
	RandomSubSpace	1,22	1,17	0,85	0,43	0,92
	RandomCommittee	1,22	1,17	0,88	0,46	0,93
sklearn.linear_model	LinearRegression(sk)	1,22	1,16	0,85	0,42	0,91
	Ridge	1,22	1,17	0,85	0,42	0,92
	BayesianRidge	1,22	1,17	0,85	0,42	0,92
sklearn.tree	DecisionTree	1,22	1,17	0,92	0,51	0,96
sklearn.neighbors	KNeighbors	1,25	1,16	0,84	0,42	0,92
xgboost	XGB	1,22	1,15	0,82	0,41	0,90
sklearn.svm	SVR(SVM)	1,07	1,09	0,81	0,42	0,85
keras	ann-dense	1,19	1,19	0,88	0,44	0,93
	ann-avg	1,27	1,22	0,95	0,54	1,00
lightgbm	LGBM	1,22	1,17	0,83	0,46	0,92

Como podemos observar en la tabla 1, cuando se incrementa el número de características, el error promedio decrece. Los algoritmos más precisos son aquellos que acumulan más resultados sombreados en verde, dado que significa que pertenecen al primer cuartil (mejores resultados). En el caso contrario se encuentran los algoritmos sombreados en rojo, que significa que pertenecen al último cuartil (peores resultados). Como cabía esperar, utilizar el histórico del alumno en la asignatura mejora significativamente la predicción de los sistemas.

Para probar los algoritmos con los datos descritos se ha creado un sistema interactivo online. Este sistema está disponible para su uso de forma pública en la dirección: <https://goo.gl/hCTbJj>. La implementación se ha realizado en *Google Colaboratory* usando un Notebook de Python, y permite seleccionar tanto el algoritmo a utilizar como la práctica a predecir. El propósito es ofrecer una versión

del sistema que pueda ser probada de forma pública y además que se pueda adaptar fácilmente a otras asignaturas.

4. CONCLUSIONES

En esta red docente se ha realizado un estudio de 24 algoritmos de inteligencia artificial, pertenecientes a diferentes categorías dentro del campo del aprendizaje automático, para la predicción de las calificaciones de los alumnos en una asignatura dada. Esta propuesta tiene como objetivo principal la detección temprana de anomalías en las calificaciones de los alumnos, para de esta forma asistir a la labor del profesor y ayudarle en el seguimiento de clases masificadas para que pueda localizar estos casos e intervenir a tiempo.

El sistema propuesto se ha evaluado usando datos de cuatro prácticas de 751 alumnos obtenidos de cuatro cursos académicos completos de una asignatura de informática introductoria. En esta experimentación se ha observado que los algoritmos tradicionales basados en regresión lineal múltiple obtienen buenos resultados cuando se utilizan pocas variables como base para la predicción. Cuando el número de variables aumenta, los algoritmos que obtienen mejores resultados son los correspondientes a la familia de los llamados máquina de vectores soporte (SVM - *Support Vector Machine*) y los recientes algoritmos basados en el aumentado de gradiente extremo (XGB - *eXtreme Gradient Boosting*).

5. TAREAS DESARROLLADAS EN LA RED

A continuación se enumeran los componentes de la red y las tareas que han desarrollado.

PARTICIPANTE DE LA RED	TAREAS QUE DESARROLLA
Antonio Javier Gallego Sánchez	Coordinación. Implementación y aplicación de la metodología. Aportación de ideas base.
Juan Ramón Rico Juan	Implementación y aplicación de la metodología. Aportación de ideas base.

Jorge Calvo Zaragoza	Aportación de ideas a la metodología.
Francisco José Castellanos Regalado	Aportación de ideas a la metodología.
David Rizo Valero	Aportación de ideas a la metodología.

6. REFERENCIAS BIBLIOGRÁFICAS

Barnes, T., Boyer, K., Sharon, I., Hsiao, H., Le, N.-T., & Sosnovsky, S. (2017) Preface for the special issue on ai-supported education in computer science. *International Journal of Artificial Intelligence in Education*, 27, 1-4.

Costa, E., Fonseca, B., Santana, M.A., de Araújo, F.F. & Rego, J. (2017) Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior*, 73:247-256.

Daud, A., Aljohani, N.R., Abbasi, R.A., Lytras, M.D., Abbas, F. & Alowibdi, J.S. (2017) Predicting student performance using advanced learning analytics. *Proceedings of the 26th International Conference on World Wide Web Companion*, 415-421.

7. REFERENCIA BIBLIOGRÁFICA DE LA PUBLICACIÓN CIENTÍFICA DE MIEMBROS DE LA RED PUBLICADA O EN PRENSA QUE COMPLEMENTA ESTA MEMORIA

Juan Ramón Rico-Juan, Antonio-Javier Gallego, Francisco J. Castellanos y Jorge Calvo-Zaragoza. Sistema para la detección temprana de anomalías en la evaluación usando técnicas de aprendizaje automático. *Actas de las Jornadas sobre Enseñanza Universitaria de la Informática*, 4 (2019), 311-318.
http://www.aenui.net/ojs/index.php?journal=actas_jenui&page=article&op=view&path%5B%5D=512