

MODELOS DE SCORING PARA DETECTAR POTENCIAL DE ABANDONO EN CARRERAS DE INGENIERÍA

Diego Edwards, Silvia N. Pérez, Mónica Giuliano, Cecilia Gargano
 Universidad Nacional de La Matanza, Argentina
 dedwards@ing.unlam.edu.ar

INTRODUCCION

El objetivo de este trabajo es mostrar resultados preliminares de un estudio para identificar factores que inciden en la deserción o abandono de carrera de los alumnos de carreras de Ingeniería de una universidad pública argentina. Se busca desarrollar un modelo de scoring para identificar el potencial de abandono en el primer año de carrera.

METODOLOGÍA

Para el estudio se analizaron los datos de alumnos de las carreras de Ingenierías Electrónica, Informática, Industrial y Civil de la Universidad Nacional de La Matanza (UNLaM), haciendo foco en la cohorte 2012. En la primera etapa, se utilizaron algoritmos de segmentación CHAID (detector de interacción Chi Cuadrado) y CART (árboles de clasificación y regresión) para distinguir las variables que mejor discriminan a la variable indicadora de abandono. Considerando dichas variables, se propuso un modelo de regresión logística buscando interpretar las relaciones halladas. Las variables seleccionadas fueron: el género del alumno, el tipo de inscripción al ingresar, si ingresó por mejor promedio del secundario o tuvo que hacer el curso de ingreso, calificación que obtuvo en el examen de admisión, y desempeño en las materias de primer año.

RESULTADOS

El ajuste logrado en la primera etapa de análisis predijo correctamente al 80% en el ambiente de validación. Algo menor fue la performance para el modelo de regresión logística, aunque este permitió mejor interpretación de los resultados. Mediante este modelo se alcanzó una precisión (Accuracy) de 0.851, considerada aceptable, aunque el valor de sensibilidad resultó de 0.6456 indicando que el modelo no resulta óptimo aún dado que sólo clasifica correctamente al 64% de los alumnos que abandonan. La especificidad fue de 0.92, indicando que los alumnos que no abandonaron fueron clasificados correctamente en un 92%. Para comparar los modelos propuestos se consideró la medida AUC, área bajo la curva ROC, considerado un indicador de la capacidad predictiva del modelo. Para el modelo elegido este valor fue de 0.81, resultando el modelo bueno para predicción. Las Figuras 1 y 2 muestran los resultados expresados anteriormente, los cuales se obtuvieron aplicando el modelo de regresión logística mediante el software estadístico R.

Confusion Matrix and Statistics

	Reference	
Prediction	0	1
0	452	56
1	41	102

Accuracy	: 0.85
Specificity	: 0.9168
Sensitivity	: 0.6456
'Positive' Class	: 1

Fig.1: Resultados del modelo de regresión logística mediante software R.

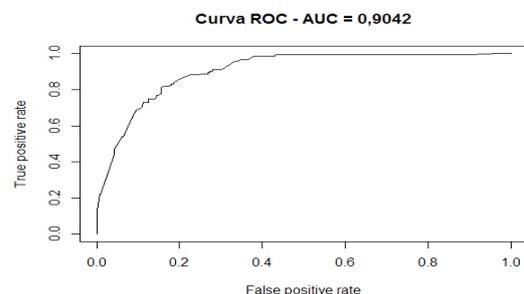


Fig.2: Curva ROC a partir de los pronósticos del modelo sugerido, con AUC = 0.81

CONCLUSIONES

Los factores de incidencia detectados en esta etapa de trabajo pueden considerarse un punto de partida para identificar y predecir grupos en riesgo de abandono en las carreras de Ingeniería de la UNLaM. Los resultados obtenidos contribuirán a mejorar las políticas institucionales para el acompañamiento y retención de los estudiantes.