UNIVERSIDAD COMPLUTENSE DE MADRID

UNIVERSIDAD
COMPLUTENSE
MADRID

OBSERVATORIO DEL ESTUDIANTE

Proyectos POE-UCM 2021

Informe final

OBSERVATORIO
DEL ESTUDIANTE
UNIVERSIDAD COMPLUTENSE DE MADRID

MODELO PREDICTIVO CON MACHINE LEARNING PARA IDENTIFICAR EL ABANDONO DE LOS ESTUDIOS UNIVERSITARIOS MEDIANTE EL ANÁLISIS DEL USO DEL CAMPUS VIRTUAL

Autores/as: Ana Belén Sánchez Prieto, Carlos Gregorio Rodríguez (directores). Josefa I Serrano Hernández, Luis Llana Díaz, María del Mar Fenoy Muñoz, José Luis Vázquez-Poletti, David Pacios Izquierdo (miembros del proyecto), Sarah Ignacio Cerrato (becaria).

Nº de proyecto: 28

Centro gestor: Facultad de Estadística

RESUMEN

El objetivo principal es analizar el abandono de los estudios universitarios enmarcado en la metodología de Big Data que integra diversas fuentes de información y herramientas informáticas además de algoritmos de Machine Learning con una perspectiva interdisciplinar.

Inicialmente el objetivo era desarrollar un modelo de Machine Learning (ML) estático a partir de los datos demográficos y socioeconómicos, y un segundo modelo de ML dinámico a partir de los datos de uso del campus virtual (CV), con la finalidad de descubrir patrones de comportamiento que indiquen que un estudiante puede estar en peligro de abandonar los estudios. Debido a la brevedad del proyecto y a distintos problemas de disponibilidad de datos solo se ha completado la primera fase, quedando pendiente la segunda.

INTRODUCCIÓN

Incluye pregunta de investigación y objetivos definidos

El punto de partida es la cuestión de si es posible identificar los patrones de comportamiento que preceden al abandono de los estudios, y si esto es posible cómo se puede ayudar a las personas susceptibles de abandonar. La cuestión se plantea como un problema de Big Data a tratar con los recursos propios de la UCM y la infraestructura de hardware y software adecuada.

Ello planteaba diversas cuestiones previas que se han resuelto satisfactoriamente (al menos hasta cierto nivel):

- Definir qué se considera realmente abandono.
- Identificar dónde se producen los abandonos.
- Establecer cuál es la causa principal de abandono.
- Averiguar qué sucede con los estudiantes tras abandonar

Quedan pendientes los siguientes objetivos:

- Identificar patrones de comportamiento en el uso del campus virtual característicos de los estudiantes en proceso de abandono.

DISEÑO METODOLÓGICO

Pensamos que es posible contestar afirmativamente a estas preguntas planteadas en los objetivos, al menos parcialmente. Además, es muy interesarte plantearlas y ver en qué medida las técnicas actuales de análisis de datos pueden utilizarse para mejorar la calidad de la docencia y la atención a los estudiantes.

Definición del concepto de abandono:

Comenzamos definiendo y clasificando los tipos de abandono (de asignaturas, de titulaciones, de estudios universitarios) y las repercusiones y costes que tienen en los distintos ámbitos: profesorado, facultades, universidad. Para ello se ha realizado un barrido de la bibliografía correspondiente.

Adquisición, integración y limpieza de datos:

La principal novedad del proyecto radica en la utilización de algoritmos de Machine Learning que utilizan las bases de datos institucionales (SIDI, GEA), después de anonimizar convenientemente los registros para respetar en todo momento las leyes vigentes sobre la protección de datos. Para ello se ha firmado un convenio con el Centro de Inteligencia Institucional de la Universidad Complutense, con cuyo director, José Arbués Bedia, se ha tenido contacto regularmente.

Por pragmatismo y para no interferir con los sistemas institucionales de la UCM, los datos procedentes de las fuentes sobredichas se fusionaron en dos bases de datos: una de las encuestas socio-económicas y la segunda con los datos académicos. Las encuestas socio-económicas resultaron ser de poco uso por haberse cambiado el formato en el año 2017 y solo existe volumen de respuestas suficiente para las preguntas relativas a la educación e ingresos familiares.

Preferentemente se han utilizado datos históricos, desde el curso 2014-2015 hasta el curso 2020-2021. Las incidencias planteadas por la pandemia de COVID-19 pueden haber causado alteraciones en los comportamientos de los estudiantes, por lo cual es necesaria una revisión dentro de dos o tres cursos. Esa misma revisión permitirá así mismo un mejor aprovechamiento de las encuestas socio-económicas, habida cuenta de que no se vuelva a cambiar el formato.

A partir de estos datos, se procedió a un análisis exploratorio, visualización preliminar y estudio preliminar de propiedades.

Entrenamiento y afinado de los distintos modelos de Machine Learning:

Se ha utilizado distintos modelos predictivos de clasificación, principalmente Random Forest, Support Vector Machine y Neural Networks.

Random Forest es el modelo que ha resultado en una mayor precisión de predicción.

RESULTADOS

Respecto de los objetivos iniciales, se han resuelto satisfactoriamente los siguientes:

- Definir qué se considera realmente abandono. La definición estandarizada define abandono cuando un estudiante no se matricula en ninguna asignatura del plan de estudios que cursa durante dos años consecutivos. Si bien esta definición es interesante para las titulaciones, no lo es para la universidad en su conjunto, pues se cuentan como abandonos cambios de plan de estudios. De forma similar, algunos abandonos son cambios de universidad y no deberían contarse como abandono universitario. Los abandonos que más nos interesa identificar son aquellos que no continúan con su formación universitaria. En especial si el abandono es debido a cuestiones socioeconómicas (falta de recursos, exclusión, etc.) que con polìticas adecuadas (becas, atención personalizada, etc.) se puédan evitar..
- Identificar dónde se producen los abandonos. Aunque el abandono se puede producir en cualquier momento, se ha detectado que la inmensa mayoría de los estudiantes que abandonan lo hacen en el primer año.
- Establecer cuál es la causa principal de abandono. Existe una fuerte correlación entre el abandono y el bajo rendimiento académico. En igualdad de condiciones en relación al rendimiento académico, los estudiantes más jóvenes son más proclives a abandonar que los más mayores.
- Averiguar qué sucede con los estudiantes tras abandonar. Aunque es imposible seguir la pista de la totalidad de los estudiantes que abandonan, muchos simplemente cambian de plan de estudios, por lo que en realidad muchos casos de aparente abandono son simplemente cambio de estudios.

CONCLUSIONES/DISCUSIÓN

El impacto de resultados debe orientarse a reducir las repercusiones adversas, tanto en lo social como en lo personal, del abandono de los estudios universitarios.

Por un lado, al disponer de una clasificación sistematizada de los factores más importantes de abandono en las distintas situaciones personales de los estudiantes vulnerables, se puede analizar para definir mejores políticas, aumentar la efectividad de los programas de becas o ayudas, en definitiva para que la UCM pueda planificar mejor.

Ello redunda en una mejora de la calidad de la docencia y en algunos de los indicadores (como la tasa de abandono) que se utilizan en las distintas evaluaciones de calidad y en los rankings de universidades.

La difusión debe orientarse primeramente en el marco de la Universidad Complutense, de modo que las distintas instancias pertinentes (Equipo Rectoral, Consejo Social, Decanatos) puedan tomar las medidas oportunas de seguimiento y apoyo de los estudiantes con más posibilidades de abandono.

Dado que el rendimiento académico durante el curso es el principal factor de abandono, es importante conciencias a los docentes de que los estudiantes más vulnerables se beneficiarían especialmente de un acompañamiento específico y una atención más personalizada.

A este respecto es necesario continuar el trabajo realizado con el previsto (pero inacabado) análisis de la utilización del campus virtual, de modo que sea posible identificar dinámicamente a los estudiantes en riesgo de abandono a través de su interacción con esta plataforma de aprendizaje. La consecuencia inmediata sería el desarrollo de una aplicación que avisase al profesor al detectar comportamientos compatibles con el abandono.

En cuanto a la difusión externa, sería interesante establecer algún tipo de colaboración con otras universidades, tanto específicamente madrileñas como españolas, con el fin de constatar si estos patrones de abandono son generales o específicamente complutenses.

PROPUESTA DE POLÍTICAS DERIVADAS DEL PROYECTO

Las acciones de futuro más directas que permitirán continuar a partir de los resultados del estudio son:

- Los modelos de predicción se pueden ir mejorando y adaptándose a medida que se vayan incorporando nuevos datos en los cursos sucesivos.
- Integración de los modelos predictivos en el campus virtual de la Universidad Complutense, de modo que las diferentes estructuras de la UCM relacionadas con estudiantes (rectorado, facultades, coordinadores, docentes, etc.) tengan alertas sobre los alumnos en riesgo de abandono.