

Sistema de Alerta Temprana de Estudiantes Universitarios con Riesgo de Abandono

Algoritmos de Machine Learning y Datos Abiertos

Proyecto presentado al concurso “II Datathon UniversiDATA”

Fernando A. López Hernández
Catedrático de Universidad
Universidad Politécnica de Cartagena

Abstract

Uno de los principales problemas del sistema educativo español es el alto porcentaje de estudiantes que abandonan la educación superior. Según el informe “Datos y Cifras del Sistema Universitario Español. Publicación 2022-2023”, uno de cada tres estudiantes universitarios deja sus estudios o cambia de titulación. Implementar un sistema de alerta temprana basado en el análisis de datos masivos puede ser una herramienta eficaz para reducir esta tasa de abandono. Los resultados de este preanálisis, con más de 200.000 estudiantes de primer curso, emplea un algoritmo de aprendizaje automático que asigna a cada estudiante una probabilidad de abandono, permitiendo identificar a aquellos con mayor riesgo. Con base en estas probabilidades, se pueden diseñar estrategias específicas para tratar a los estudiantes según sus características individuales. Esta metodología no solo ayuda a disminuir el abandono, sino que también optimiza los recursos educativos y mejora la experiencia de los estudiantes y de sus familias.

Introducción

Título del proyecto

Sistema de Alerta Temprana de Estudiantes Universitarios con Riesgo de Abandono. Algoritmos de Machine Learning y Datos Abiertos.

Problemática a resolver

Detalle el problema o desafío que quiere resolver o ayudar a resolver con su proyecto

Uno de los grandes problemas del sistema educativo español es el alto porcentaje de estudiantes que abandonan la educación superior. Según el informe “Datos y Cifras del Sistema

Universitario Español. Publicación 2022-2023” uno de cada tres alumnos universitarios abandona sus estudios o cambia de titulación (Tabla 5.4.1). La mayoría, un 22%, abandona el primer año (Tabla 5.4.6).

No se trata de algo coyuntural sino que es un problema persistente en el sistema educativo español, también en el europeo (Kehm, Larsen, and Sommersel 2019), que afecta no solo a la trayectoria académica de los estudiantes, sino también a la economía y el desarrollo social del país. Según datos recientes, un porcentaje significativo de estudiantes abandona sus estudios superiores antes de finalizar su titulación, lo que refleja una inversión educativa no aprovechada y una pérdida de talento potencial para la sociedad.

Las causas del abandono son diversas y complejas (Constante-Amores et al. 2021), incluyendo factores académicos, económicos, sociales y personales. Entre los motivos más comunes se encuentran la falta de motivación, problemas financieros, la elección incorrecta de la carrera, y dificultades en la adaptación al entorno universitario. Este problema no solo representa un desafío para los estudiantes, sino que también pone en cuestión la efectividad de las políticas educativas y la capacidad de las universidades para retener a sus alumnos. Abordar este problema requiere de estrategias integrales que incluyan la **identificación temprana de estudiantes en riesgo**, el fortalecimiento del apoyo académico y personal, y la implementación de políticas que faciliten la permanencia y el éxito académico de todos los estudiantes.

En el ámbito empresarial el abandono de clientes (término inglés *churn*) es una cuestión crucial que la mayoría de las grandes empresas abordan basándose en el análisis cuantitativos, dedicando grandes esfuerzos a la retención de clientes. El coste de incorporar a un nuevo cliente es siete veces superior al de retenerlo (Llave Montiel and Lopez 2020). Los clientes leales tienden a realizar mayores gastos en la compañía, recomiendan el negocio a otros y proporcionan un flujo de ingresos constante. Además, una alta retención refleja satisfacción y confianza en la empresa, fortaleciendo su reputación y competitividad en el mercado. En esa misma línea debería de trabajar la universidad española para la retención de estudiantes.

Solución propuesta

Detalle cómo sería la solución que propone, tanto desde un punto de vista funcional (perspectiva de los usuarios: ¿cómo obtendrán los resultados? ¿cómo y para qué los usarán? ¿cómo interactuarán con el producto?) como técnico (perspectiva de los desarrolladores: ¿cuáles serán los componentes y tecnologías implicadas?)

Las universidades disponen de una ingente cantidad de información que permite desarrollar modelos predictivos para identificar a los estudiantes con riesgo de abandono. Asignar una probabilidad de abandono a cada estudiante en diferentes etapas del inicio de los estudios universitarios permitirá a las instituciones implementar estrategias para reducir las tasas de deserción.

Hay tres periodos clave para evaluar el riesgo de abandono:

- (i) antes del inicio del primer curso académico, mediante la elaboración de un modelo que incorpore solo información correspondiente a los datos de acceso del alumno.
- (ii) al finalizar el primer cuatrimestre del primer curso, utilizando tanto información correspondiente a los datos de acceso del alumno como a los resultados obtenidos en el primer cuatrimestre, y
- (iii) al finalizar el primer curso académico.

Además de los programas clásicos de apoyo académico personalizado, este proyecto propone agrupar a los estudiantes en función de sus probabilidades de abandono. La asignación de un estudiante a un determinado grupo (A, B, C, etc) no debe ser aleatoria, sino que debe basarse en la similitud de sus probabilidades de abandono con el objetivo de proporcionarles el apoyo adecuado dependiendo del perfil del estudiante.

Por ejemplo, estudiantes con mayor probabilidad de abandono pueden ser asignados a grupos con profesores específicos que se especialicen en retención estudiantil (p.e. profesores con mejores valoraciones en las encuestas docentes o con perfiles más similares a profesores de bachiller - profesores asociados que den clase en institutos - o profesores especialmente formados para este fin). Este proceso de asignación debe llevarse a cabo en cada uno de los periodos mencionados (i, ii, iii) evaluando en cada periodo la probabilidad de abandono y reasignando estudiantes a uno u otro grupo para maximizar la efectividad de las intervenciones. Esta reasignación de los estudiantes para maximizar el impacto de las intervenciones puede implique un cambio en su entorno social académico.

Identificación de destinatarios

¿A quién le sería de utilidad el proyecto? ¿Quiénes serían los destinatarios de los resultados? ¿Qué personas o colectivos podrían estar interesados en obtenerlos y para qué les servirían?

Un proyecto centrado en la reducción de la tasa de abandono en la universidad sería de utilidad para una varios actores. En primer lugar, **las universidades** se beneficiarían al mejorar sus tasas de retención, lo cual influirá positivamente en su reputación, rankings académicos, y financiación. En segundo lugar, **los estudiantes** también serían beneficiarios directos, ya que un proyecto así podría ofrecerles un apoyo más personalizado, aumentando sus posibilidades de completar sus estudios y reducir el riesgo de abandono por problemas académicos. En tercer lugar, se beneficiarían **las familias de los estudiantes** al ver un mejor retorno de su inversión en la educación superior. Por último, **la sociedad en general** podría beneficiarse de un mayor número de graduados cualificados, lo que contribuiría al desarrollo socioeconómico y a una mejor preparación de la fuerza laboral para enfrentar los desafíos del mercado laboral.

Utilidad del proyecto

¿Cómo les servirían los resultados del proyecto a los destinatarios para resolver la problemática identificada? ¿De qué forma mejoraría su vida/trabajo? ¿Tiene

el proyecto impacto indirecto o secundario en algún otro ámbito, además de en los destinatarios directos?

Los resultados de un proyecto centrado en la elaboración de un sistema de detección temprana del abandono ayudará tanto a los estudiantes como a las universidades. Para las **universidades**, estos resultados permitirían desarrollar estrategias más efectivas para identificar a los estudiantes en riesgo y proporcionarles apoyo personalizado, mejorando así la retención y, por ende, la estabilidad financiera y la reputación académica de la institución. Para los **estudiantes**, recibir un apoyo adecuado aumentaría sus oportunidades de éxito académico y personal, mejorando sus perspectivas laborales y la satisfacción con su experiencia universitaria. El proyecto podría tener un impacto indirecto en otros ámbitos, como el mercado laboral, al aumentar la cantidad de graduados cualificados, y en las familias, que verían una mejor inversión en la educación de sus hijos. También podría influir en las políticas educativas, promoviendo mejoras sistémicas basadas en los hallazgos del proyecto.

Conjuntos de datos a utilizar

Enumere los datasets de UniversiDATA u otras fuentes externas que tiene previsto utilizar en el proyecto. Es importante haber hecho un preanálisis de los mismos para confirmar que dan cobertura al uso que quiere darles en el proyecto. SE VALORARÁ ESPECIALMENTE QUE SE RECOJAN EN ESTE APARTADO LAS CONCLUSIONES DEL PREANÁLISIS DE LOS DATASETS IMPLICADOS, ATENDIENDO ESPECIALMENTE AL IMPACTO DE LOS PROCESOS DE ANONIMIZACIÓN.

El ejercicio que se propone en este proyecto está basado en la aplicación de un algoritmo de aprendizaje automático a las bases de datos disponibles en abierto en el portal UniversiDATA. Debido al fuerte proceso de anonimización de esta bases de datos, los resultados que se obtienen están lejos de tener la capacidad predictiva que se obtendría con una base de datos sin anonimizar pero servirá de ejemplo para ilustrar la idea que subyace en este proyecto.

La principal base de datos utilizada en este proyecto es “Matrículas” (todos los cursos disponibles). Los datos correspondientes al bloque “Campos pivote” junto con el bloque de coherencia “Bloque 11” que suministra información sobre los resultados académicos del estudiante servirán para identificar alumnos matriculados en primer curso cuyo rendimiento académico haya sido bajo. Estudiantes con bajo rendimiento académico¹ durante el primer año serán identificados como estudiantes que abandonan sus estudios. La literatura avala esta hipótesis: **Los datos de rendimiento académico son un buen predictor del abandono** (Ortiz-Lozano et al. 2020).

Además de la base de datos “Matrículas” también se utilizará la base de datos “Títulos” para identificar la rama de conocimiento de cada titulación, la base de “Acceso” para conocer características socio-demográficas de los estudiantes y la base de datos “Personal-PDI” (sección “Recursos Humanos”). Debido al proceso de anonimización el nexo de unión entre todas las bases de datos será la titulación (“cod_titulacion”), asignando a cada uno de los

¹Tasa rendimiento = Número de créditos superados entre total créditos matriculados en el curso

estudiantes matriculados un indicador basado en los datos globales de la titulación que curse. Por ejemplo, a cada estudiante de la base de datos “Matrículas” que curse estudios en la titulación “X” se le asignará el porcentaje de estudiantes de esa misma titulación cuyo padre tenga estudios universitarios. En el caso de la base “Personal-PDI” el nexo de unión será la universidad. En el caso de no existir proceso de anonimización se imputarían estos indicadores a cada individuo.

Se ha seleccionado el algoritmo de aprendizaje automático *Multivariate Adaptive Regression Splines* (MARS), aplicado a un modelo de regresión logística, para desarrollar un sistema de alerta temprana que identifique el riesgo de abandono de cada estudiante. A diferencia de otros algoritmos similares, MARS tiene la ventaja de seleccionar automáticamente las variables más relevantes que determinan las probabilidades de abandono, considerando también impactos no lineales de las variables independientes. Una de las principales características de este algoritmo es su fácil interpretación. Esta facilidad de interpretación es crucial, ya que permite a las universidades entender claramente los factores que influyen en el abandono y, por tanto, tomar decisiones rápidas y efectivas para reagrupar a los estudiantes en función de su riesgo.

Preanálisis

Toda la investigación se ha llevado a cabo siguiendo estrictamente el principio de reproducibilidad, lo que garantiza que los resultados puedan ser verificados y replicados por otros investigadores. Para cumplir con este estándar, se proporciona acceso completo a los datos abiertos utilizados en la investigación y se ha puesto a disposición el código empleado en el análisis en formato RMarkdown. Todo este material está accesible en nuestro repositorio de GitHub, lo que facilita no solo la transparencia y la validación independiente de los resultados, sino también la posibilidad de que otros puedan reutilizar y adaptar estos recursos para investigaciones futuras o para mejorar metodologías existentes.

<https://github.com/f8l5h9/UniversiDATA>

Se ha realizado un preanálisis de los datos correspondiente al periodo (iii) que puede servir de ejemplo de cómo desarrollar un sistema de alerta temprana en los periodos (i) y (ii).

En primer lugar se ha identificado al alumno/a de primero en la base de datos “Matrículas” como aquel estudiante que cumple varias características (“num_creditos_mat_1_curso==num_total_creditos_mat_1_curso”; “ind_se_titula_curso==“NO”; “num_total_creditos_mat_curso!=0”; ...). En segundo lugar, se combina la información de las bases de datos “Titulación”, “Acceso” y “Presupuesto de Gastos” para añadir a cada individuo los datos correspondientes a la titulación en la que está matriculado. En tercer lugar, se clasifica a los estudiantes en las categorías de “abandono” y “no abandono” según su tasa de rendimiento académico, considerando como abandono una tasa inferior al 30% (aunque se analizarán diferentes umbrales). Finalmente, se aplicará el algoritmo MARS a un modelo de regresión logística, utilizando toda la información disponible en estas bases de datos.

En el preanálisis realizado para lanzar esta propuesta, con más de 200.000 observaciones (ver código en GitHub), el algoritmo MARS identifica algunas variables clave que determinan la pertenencia a la categoría “abandono”. Se muestran como variables relevantes el género,

matricularse en una doble titulación, pertenencia a la rama de Arquitectura e Ingeniería, pertenencia a un centro adscrito, etc. También el algoritmo considera como relevantes la interacción entre algunas de las variables incluidas en el preanálisis.

Un análisis completo se realizará en el caso de que el proyecto sea seleccionado. El código reproducible correspondiente a este preanálisis está disponible en el cuaderno de RMarkdown titulado “UniversiDATA.Rmd”.

Código reproducible

```
library(stringr)
library(openxlsx)
library(readxl)
library(dplyr)
library(earth)
library(plotmo)
```

Datos de Matrículas

Lectura de las bases de datos “Matriculas”. Selección solo alumnos de grado y asignación de la edad del alumno

```
# UCM
# =====
ucm.1718 <- read_xlsx("Data/ucm-matriculas-2017-18-anonimizado_4.xlsx")
ucm.1718$edad <- 2018-ucm.1718$anio_nacimiento
ucm.1819 <- read_xlsx("Data/ucm-matriculas-2018-19-anonimizado_4.xlsx")
ucm.1819$edad <- 2019-ucm.1819$anio_nacimiento
ucm.1920 <- read_xlsx("Data/ucm-matriculas-2019-20-anonimizado.xlsx")
ucm.1920$edad <- 2020-ucm.1920$anio_nacimiento
ucm.2021 <- read_xlsx("Data/ucm-matriculas-2020-21-anonimizado.xlsx")
ucm.2021$edad <- 2021-ucm.2021$anio_nacimiento
ucm.2122 <- read_xlsx("Data/ucm-matriculas-2021-22-anonimizado.xlsx")
ucm.2122$edad <- 2022-ucm.2122$anio_nacimiento
ucm.2223 <- read_xlsx("Data/ucm-matriculas-2022-23-anonimizado.xlsx")
ucm.2223$edad <- 2023-ucm.2223$anio_nacimiento
ucm <- rbind(ucm.1718,ucm.1819,ucm.1920,ucm.2021,ucm.2122,ucm.2223)
rm(ucm.1718,ucm.1819,ucm.1920,ucm.2021,ucm.2122,ucm.2223)
ucm <- ucm[ucm$des_tipo_estudio=="Grado",]

# UAM
# =====
uam.1718 <- read_xlsx("Data/uam-matriculas-2017-18-anonimizado_2.xlsx")
uam.1718$edad <- 2018-uam.1718$anio_nacimiento
```

```

uam.1819 <- read_xlsx("Data/uam-matriculas-2018-19-anonimizado_1.xlsx")
uam.1819$edad <- 2019-uam.1819$anio_nacimiento
uam.1920 <- read_xlsx("Data/uam-matriculas-2019-20-anonimizado.xlsx")
uam.1920$edad <- 2020-uam.1920$anio_nacimiento
uam.2021 <- read_xlsx("Data/uam-matriculas-2020-21-anonimizado.xlsx")
uam.2021$edad <- 2021-uam.2021$anio_nacimiento
uam.2122 <- read_xlsx("Data/uam-matriculas-2021-22-anonimizado.xlsx")
uam.2122$edad <- 2022-uam.2122$anio_nacimiento
uam.2223 <- read_xlsx("Data/uam-matriculas-2022-23-anonimizado.xlsx")
uam.2223$edad <- 2023-uam.2223$anio_nacimiento
uam <- rbind(uam.1718,uam.1819,uam.1920,uam.2021,uam.2122,uam.2223)
rm(uam.1718,uam.1819,uam.1920,uam.2021,uam.2122,uam.2223)
uam <- uam[uam$des_tipo_estudio=="Grado",]

# UVa
# =====
uva.1718 <- read_xlsx("Data/uva-matriculas-2017-18-anonimizado_1.xlsx")
uva.1718$edad <- 2018-uva.1718$anio_nacimiento
uva.1819 <- read_xlsx("Data/uva-matriculas-2018-19-anonimizado_1.xlsx")
uva.1819$edad <- 2019-uva.1819$anio_nacimiento
uva.1920 <- read_xlsx("Data/uva-matriculas-2019-20-anonimizado_1.xlsx")
uva.1920$edad <- 2020-uva.1920$anio_nacimiento
uva.2021 <- read_xlsx("Data/uva-matriculas-2020-21-anonimizado.xlsx")
uva.2021$edad <- 2021-uva.2021$anio_nacimiento
uva.2122 <- read_xlsx("Data/uva-matriculas-2021-22-anonimizado.xlsx")
uva.2122$edad <- 2022-uva.2122$anio_nacimiento
uva.2223 <- read_xlsx("Data/uva-matriculas-2022-23-anonimizado.xlsx")
uva.2223$edad <- 2023-uva.2223$anio_nacimiento
uva <- rbind(uva.1718,uva.1819,uva.1920,uva.2021,uva.2122,uva.2223)
rm(uva.1718,uva.1819,uva.1920,uva.2021,uva.2122,uva.2223)
uva <- uva[uva$des_tipo_estudio=="Grado",]

# UC3M
# =====
uc3m.1718 <- read_xlsx("Data/uc3m-matriculas-2017-18-anonimizado_0.xlsx")
uc3m.1718$edad <- 2018-uc3m.1718$anio_nacimiento
uc3m.1819 <- read_xlsx("Data/uc3m-matriculas-2018-19-anonimizado_0.xlsx")
uc3m.1819$edad <- 2019-uc3m.1819$anio_nacimiento
uc3m.1920 <- read_xlsx("Data/uc3m-matriculas-2019-20-anonimizado_0.xlsx")
uc3m.1920$edad <- 2020-uc3m.1920$anio_nacimiento
uc3m.2021 <- read_xlsx("Data/uc3m-matriculas-2020-21-anonimizado.xlsx")
uc3m.2021$edad <- 2021-uc3m.2021$anio_nacimiento
uc3m.2122 <- read_xlsx("Data/uc3m-matriculas-2021-22-anonimizado.xlsx")
uc3m.2122$edad <- 2022-uc3m.2122$anio_nacimiento

```

```

uc3m.2223 <- read_xlsx("Data/uc3m-matriculas-2022-23-anonimizado.xlsx")
uc3m.2223$edad <- 2023-uc3m.2223$anio_nacimiento
uc3m <- rbind(uc3m.1718,uc3m.1819,uc3m.1920,uc3m.2021,uc3m.2122,uc3m.2223)
rm(uc3m.1718,uc3m.1819,uc3m.1920,uc3m.2021,uc3m.2122,uc3m.2223)
uc3m <- uc3m[uc3m$des_tipo_estudio=="Grado",]

# URJC
# =====
urjc.1718 <- read_xlsx("Data/urjc-matriculas-2017-18-anonimizado_2.xlsx")
urjc.1718$edad <- 2018-urjc.1718$anio_nacimiento
urjc.1819 <- read_xlsx("Data/urjc-matriculas-2018-19-anonimizado_1.xlsx")
urjc.1819$edad <- 2019-urjc.1819$anio_nacimiento
urjc.1920 <- read_xlsx("Data/urjc-matriculas-2019-20-anonimizado.xlsx")
urjc.1920$edad <- 2020-urjc.1920$anio_nacimiento
urjc.2021 <- read_xlsx("Data/urjc-matriculas-2020-21-anonimizado.xlsx")
urjc.2021$edad <- 2021-urjc.2021$anio_nacimiento
urjc.2122 <- read_xlsx("Data/urjc-matriculas-2021-22-anonimizado.xlsx")
urjc.2122$edad <- 2022-urjc.2122$anio_nacimiento
urjc.2223 <- read_xlsx("Data/urjc-matriculas-2022-23-anonimizado.xlsx")
urjc.2223$edad <- 2023-urjc.2223$anio_nacimiento
urjc <- rbind(urjc.1718,urjc.1819,urjc.1920,urjc.2021,urjc.2122,urjc.2223)
rm(urjc.1718,urjc.1819,urjc.1920,urjc.2021,urjc.2122,urjc.2223)
urjc <- urjc[urjc$des_tipo_estudio=="Grado",]

```

Filtro alumnos de primer curso, primera matrícula

Selección solo de alumnos de primer curso imponiendo restricciones sobre las variables asociadas al número de créditos (Bloque 11)

```

# Alumnos de primero
# UCM
# UCM =====
bd <- ucm
bd <- bd[!is.na(bd$des_titulacion),]
bd <- bd[bd$ind_se_titula_curso=="NO",]
bd <- bd[bd$num_total_creditos_mat_curso!=0,]
bd <- bd[bd$num_total_creditos_transf_inicio==0,]
bd <- bd[bd$num_total_creditos_rec_conv_inicio==0,]
bd <- bd[(bd$num_creditos_mat_1_curso==bd$num_total_creditos_mat_inicio),]
bd <- bd[,c(2,4,26,28,42,43,47,51:63,81)]
bd$adscrito <- as.factor(!str_detect(bd$des_centro,"Facultad"))
bd$des_genero <- as.factor(bd$des_genero)
bd$curso_academico <- as.factor(bd$curso_academico)
bd$cod_universidad <- as.factor(bd$cod_universidad)
bd$doble <- as.factor(str_detect(bd$des_titulacion,"DOBLE"))

```



```

bd$mm <- bd$des_municipio_residencia=="MADRID"
bd$pm <- bd$des_provincia_residencia=="Madrid"
hh <- bd %>% group_by(cod_titulacion) %>%
  summarise(n.titula=n(),
            edad.media.tit = mean(edad,na.rm=T),
            muni.local=sum(mm,na.rm=T)/n.titula,
            prov.local=sum(pm,na.rm=T)/n.titula)
bd <- merge(bd,hh,by="cod_titulacion",all.x = TRUE)
bd.ucm <- bd

# UAM
# =====
bd <- uam
bd <- bd[!is.na(bd$des_titulacion),]
bd <- bd[bd$num_total_creditos_mat_curso!=0,]
bd <- bd[bd$ind_se_titula_curso=="NO",]
bd <- bd[bd$num_total_creditos_transf_inicio==0,]
bd <- bd[bd$num_total_creditos_rec_conv_inicio==0,]
bd <- bd[(bd$num_creditos_mat_1_curso==bd$num_total_creditos_mat_inicio),]
bd <- bd[,c(2,4,26,28,42,43,47,51:63,81)]
bd$adscrito <- as.factor(!str_detect(bd$des_centro,"Facultad"))
bd$des_genero <- as.factor(bd$des_genero)
bd$curso_academico <- as.factor(bd$curso_academico)
bd$cod_universidad <- as.factor(bd$cod_universidad)
bd$doble <- as.factor(str_detect(bd$des_titulacion," / "))
bd$mm <- bd$des_municipio_residencia=="MADRID"
bd$pm <- bd$des_provincia_residencia=="Madrid"
hh <- bd %>% group_by(cod_titulacion) %>%
  summarise(n.titula=n(),
            edad.media.tit = mean(edad,na.rm=T),
            muni.local=sum(mm,na.rm=T)/n.titula,
            prov.local=sum(pm,na.rm=T)/n.titula)
bd <- merge(bd,hh,by="cod_titulacion",all.x = TRUE)
bd.uam <- bd

# Alumnos de primero
# UVa
# =====
bd <- uva
bd <- bd[!is.na(bd$des_titulacion),]
bd <- bd[bd$num_total_creditos_mat_curso!=0,]
bd <- bd[bd$ind_se_titula_curso=="NO",]
bd <- bd[bd$num_total_creditos_transf_inicio==0,]
bd <- bd[bd$num_total_creditos_rec_conv_inicio==0,]

```

```

bd <- bd[(bd$num_creditos_mat_1_curso==bd$num_total_creditos_mat_inicio),]
bd <- bd[,c(2,4,26,28,42,43,47,51:63,81)]
bd$adscrito <- as.factor(str_detect(bd$des_centro,c("ADSCRITA|Adscrita")))
bd$des_genero <- as.factor(bd$des_genero)
bd$curso_academico <- as.factor(bd$curso_academico)
bd$cod_universidad <- as.factor(bd$cod_universidad)
bd$doble <- as.factor(str_detect(bd$des_titulacion,"Doble|conjunto"))
bd$mm <- bd$des_municipio_residencia=="VALLADOLID"
bd$pm <- bd$des_provincia_residencia=="Valladolid"
hh <- bd %>% group_by(cod_titulacion) %>%
  summarise(n.titula=n(),
            edad.media.tit = mean(edad,na.rm=T),
            muni.local=sum(mm,na.rm=T)/n.titula,
            prov.local=sum(pm,na.rm=T)/n.titula)
bd <- merge(bd,hh,by="cod_titulacion",all.x = TRUE)
bd.uva <- bd

# Alumnos de primero
# UC3M
# =====
bd <- uc3m
bd <- bd[!is.na(bd$des_titulacion),]
bd <- bd[bd$num_total_creditos_mat_curso!=0,]
bd <- bd[bd$ind_se_titula_curso=="NO",]
bd <- bd[bd$num_total_creditos_transf_inicio==0,]
bd <- bd[bd$num_total_creditos_rec_conv_inicio==0,]
bd <- bd[(bd$num_creditos_mat_1_curso==bd$num_total_creditos_mat_inicio),]
bd <- bd[,c(2,4,26,28,42,43,47,51:63,81)]
bd$adscrito <- as.factor(str_detect(bd$des_centro,c("Guardia Civil")))
bd$des_genero <- as.factor(bd$des_genero)
bd$curso_academico <- as.factor(bd$curso_academico)
bd$cod_universidad <- as.factor(bd$cod_universidad)
bd$doble <- as.factor(str_detect(bd$des_titulacion,"Doble"))
bd$mm <- bd$des_municipio_residencia=="MADRID"
bd$pm <- bd$des_provincia_residencia=="Madrid"
hh <- bd %>% group_by(cod_titulacion) %>%
  summarise(n.titula=n(),
            edad.media.tit = mean(edad,na.rm=T),
            muni.local=sum(mm,na.rm=T)/n.titula,
            prov.local=sum(pm,na.rm=T)/n.titula)
bd <- merge(bd,hh,by="cod_titulacion",all.x = TRUE)
bd.uc3m <- bd

# Alumnos de primero

```

```

# URJC
# =====
bd <- urjc
bd <- bd[!is.na(bd$des_titulacion),]
bd <- bd[bd$num_total_creditos_mat_curso!=0,]
bd <- bd[bd$ind_se_titula_curso=="NO",]
bd <- bd[bd$num_total_creditos_transf_inicio==0,]
bd <- bd[bd$num_total_creditos_rec_conv_inicio==0,]
bd <- bd[(bd$num_creditos_mat_1_curso==bd$num_total_creditos_mat_inicio),]
bd <- bd[,c(2,4,26,28,42,43,47,51:63,81)]
##
bd$adscrito <- as.factor(str_detect(bd$des_centro,c("CEDEU|IEB|Espectaculos TAI|Escuela
bd$des_genero <- as.factor(bd$des_genero)
bd$curso_academico <- as.factor(bd$curso_academico)
bd$cod_universidad <- as.factor(bd$cod_universidad)
bd$doble <- as.factor(str_detect(bd$des_titulacion,"DOBLE"))
bd$mm <- bd$des_municipio_residencia=="MADRID"
bd$pm <- bd$des_provincia_residencia=="Madrid"
hh <- bd %>% group_by(cod_titulacion) %>%
  summarise(n.titula=n(),
            edad.media.tit = mean(edad,na.rm=T),
            muni.local=sum(mm,na.rm=T)/n.titula,
            prov.local=sum(pm,na.rm=T)/n.titula)
bd <- merge(bd,hh,by="cod_titulacion",all.x = TRUE)
bd.urjc <- bd
rm(bd,hh)

```

Rama de conocimiento

Identificación de la rama de conocimiento asociada a cada titulación

```

# UCM
# =====
ucm.t.1718 <- read.xlsx("Data/ucm-titulaciones-2017-18_0.xlsx")
ucm.t.1920 <- read.xlsx("Data/ucm-titulaciones-2019-20_1.xlsx")
ucm.t.2122 <- read.xlsx("Data/ucm-titulaciones-2021-22.xlsx")
ucm.t.2324 <- read.xlsx("Data/ucm-titulaciones-2023-24.xlsx")
ucm.t <- rbind(ucm.t.1718,ucm.t.1920,ucm.t.2122,ucm.t.2324)
rm(ucm.t.1718,ucm.t.1920,ucm.t.2122,ucm.t.2324)
ucm.t <- ucm.t[ucm.t$des_tipo_estudio=="Grado",]
ucm.t <- ucm.t[,c(3,10,12,13)]
ucm.t <- ucm.t[!duplicated(ucm.t$cod_titulacion),]
ucm.t <- ucm.t[!(ucm.t$cod_titulacion==7000815 & ucm.t$cod_rama==3),]
bd.ucm <- merge(bd.ucm,ucm.t,by="cod_titulacion")
bd.ucm$des_rama <- as.factor(bd.ucm$des_rama)

```

```

# UAM
# =====
uam.t.1718 <- read.xlsx("Data/uam-titulaciones-2017-18_1.xlsx")
uam.t.1920 <- read.xlsx("Data/uam-titulaciones-2019-20_1.xlsx")
uam.t.2122 <- read.xlsx("Data/uam-titulaciones-2021-22.xlsx")
uam.t.2324 <- read.xlsx("Data/uam-titulaciones-2023-24.xlsx")
uam.t <- rbind(uam.t.1718,uam.t.1920,uam.t.2122,uam.t.2324)
rm(uam.t.1718,uam.t.1920,uam.t.2122,uam.t.2324)
uam.t <- uam.t[uam.t$des_tipo_estudio=="Grado",]
uam.t <- uam.t[,c(3,10,12,13)]
uam.t <- uam.t[!duplicated(uam.t$cod_titulacion),]
uam.t <- uam.t[(uam.t$cod_titulacion==2502531 & uam.t$cod_rama==1),]
uam.t <- uam.t[(uam.t$cod_titulacion==2502958 & uam.t$cod_rama==1),]
bd.uam <- merge(bd.uam,uam.t,by="cod_titulacion")
bd.uam$des_rama <- as.factor(bd.uam$des_rama)

# UVA
# =====
uva.t.1718 <- read.xlsx("Data/uva-titulaciones-2017-18_0.xlsx")
uva.t.1920 <- read.xlsx("Data/uva-titulaciones-2019-20_0.xlsx")
uva.t.2122 <- read.xlsx("Data/uva-titulaciones-2021-22.xlsx")
uva.t.2223 <- read.xlsx("Data/uva-titulaciones-2022-23.xlsx")
uva.t <- rbind(uva.t.1718,uva.t.1920,uva.t.2122,uva.t.2223)
rm(uva.t.1718,uva.t.1920,uva.t.2122,uva.t.2223)
uva.t <- uva.t[uva.t$des_tipo_estudio=="Grado",]
uva.t <- uva.t[,c(3,10,12,13)]
uva.t <- uva.t[!duplicated(uva.t$cod_titulacion),]
bd.uva <- merge(bd.uva,uva.t,by="cod_titulacion")
bd.uva$des_rama <- as.factor(bd.uva$des_rama)

# UC3M
# =====
uc3m.t.1718 <- read.xlsx("Data/uc3m-titulaciones-2017-18_0.xlsx")
uc3m.t.1920 <- read.xlsx("Data/uc3m-titulaciones-2019-20_0.xlsx")
uc3m.t.2122 <- read.xlsx("Data/uc3m-titulaciones-2021-22.xlsx")
uc3m.t.2223 <- read.xlsx("Data/uc3m-titulaciones-2022-23.xlsx")
uc3m.t <- rbind(uc3m.t.1718,uc3m.t.1920,uc3m.t.2122,uc3m.t.2223)
rm(uc3m.t.1718,uc3m.t.1920,uc3m.t.2122,uc3m.t.2223)
uc3m.t <- uc3m.t[uc3m.t$des_tipo_estudio=="Grado",]
uc3m.t <- uc3m.t[,c(3,10,12,13)]
uc3m.t <- uc3m.t[!duplicated(uc3m.t$cod_titulacion),]
bd.uc3m <- merge(bd.uc3m,uc3m.t,by="cod_titulacion")
bd.uc3m$des_rama <- as.factor(bd.uc3m$des_rama)

```

```

# URJC
# =====
urjc.t.1718 <- read.xlsx("Data/urjc-titulaciones-2017-18_2.xlsx")
urjc.t.1920 <- read.xlsx("Data/urjc-titulaciones-2019-20_2.xlsx")
urjc.t.2122 <- read.xlsx("Data/urjc-titulaciones-2021-22_1.xlsx")
urjc.t.2223 <- read.xlsx("Data/urjc-titulaciones-2022-23.xlsx")
urjc.t <- rbind(urjc.t.1718,urjc.t.1920,urjc.t.2122,urjc.t.2223)
rm(urjc.t.1718,urjc.t.1920,urjc.t.2122,urjc.t.2223)
urjc.t <- urjc.t[urjc.t$des_tipo_estudio=="Grado",]
urjc.t <- urjc.t[,c(3,10,12,13)]
urjc.t <- unique(urjc.t)
urjc.t <- urjc.t[!duplicated(urjc.t$cod_titulacion),]
bd.urjc <- merge(bd.urjc,urjc.t,by="cod_titulacion")
bd.urjc$des_rama <- as.factor(bd.urjc$des_rama)
#
bd.t <- rbind(ucm.t,uam.t,uc3m.t,urjc.t,uva.t)
rm(ucm.t,uam.t,uc3m.t,urjc.t,uva.t)
bd.t <- bd.t[!duplicated(bd.t$cod_titulacion), ]

```

Datos de Acceso

Lectura de base de datos de “Acceso” e incorporación de indicadores generales a cada estudiante de primero badado en la titulación en la que está matriculado (“cod_titulacion”).

```

# UCM
# =====
ucm.a.1718 <- read.xlsx("Data/ucm-acceso-2017-18-anonimizado_0.xlsx")
ucm.a.1819 <- read.xlsx("Data/ucm-acceso-2018-19-anonimizado_0.xlsx")
ucm.a.1920 <- read.xlsx("Data/ucm-acceso-2019-20-anonimizado_0.xlsx")
ucm.a.2021 <- read.xlsx("Data/ucm-acceso-2020-21-anonimizado_0.xlsx")
ucm.a.2122 <- read.xlsx("Data/ucm-acceso-2021-22-anonimizado_0.xlsx")
ucm.a.2223 <- read.xlsx("Data/ucm-acceso-2022-23-anonimizado.xlsx")
ucm.a <- rbind(ucm.a.1718,ucm.a.1819,ucm.a.1920,ucm.a.2021,ucm.a.2122,ucm.a.2223)
rm(ucm.a.1718,ucm.a.1819,ucm.a.1920,ucm.a.2021,ucm.a.2122,ucm.a.2223)
ucm.a$uni.madre <- (ucm.a$des_nivel_estudios_madre=="Estudios Superiores")
ucm.a$uni.padre <- (ucm.a$des_nivel_estudios_padre=="Estudios Superiores")
hh <- ucm.a %>% group_by(cod_titulacion) %>%
  summarize(n.acceso=n(),
            nota.mediana=median(nota_admision,na.rm=TRUE),
            nota.min=min(nota_admision,na.rm=TRUE),
            uni.madre=sum(uni.madre,na.rm=TRUE)/n.acceso,
            uni.padre=sum(uni.padre,na.rm=TRUE)/n.acceso)
bd.ucm <- merge(bd.ucm,hh,by="cod_titulacion",all.x = TRUE)

```

```

# UAM
# =====
uam.a.1718 <- read_xlsx("Data/uam-acceso-2017-18-anonimizado_0.xlsx")
uam.a.1819 <- read_xlsx("Data/uam-acceso-2018-19-anonimizado_0.xlsx")
uam.a.1920 <- read_xlsx("Data/uam-acceso-2019-20-anonimizado_0.xlsx")
uam.a.2021 <- read_xlsx("Data/uam-acceso-2020-21-anonimizado_0.xlsx")
uam.a.2122 <- read_xlsx("Data/uam-acceso-2021-22-anonimizado_0.xlsx")
uam.a.2223 <- read_xlsx("Data/uam-acceso-2022-23-anonimizado.xlsx")
uam.a <- rbind(uam.a.1718,uam.a.1819,uam.a.1920,uam.a.2021,uam.a.2122,uam.a.2223)
rm(uam.a.1718,uam.a.1819,uam.a.1920,uam.a.2021,uam.a.2122,uam.a.2223)
uam.a$uni.madre <- (uam.a$des_nivel_estudios_madre=="Estudios Superiores")
uam.a$uni.padre <- (uam.a$des_nivel_estudios_padre=="Estudios Superiores")
hh <- uam.a %>% group_by(cod_titulacion) %>%
  summarize(n.acceso=n(),
            nota.mediana=median(nota_admision,na.rm=TRUE),
            nota.min=min(nota_admision,na.rm=TRUE),
            uni.madre=sum(uni.madre,na.rm=TRUE)/n.acceso,
            uni.padre=sum(uni.padre,na.rm=TRUE)/n.acceso)
bd.uam <- merge(bd.uam,hh,by="cod_titulacion",all.x = TRUE)

# UVa
# =====
uva.a.1718 <- read_xlsx("Data/uva-acceso-2017-18-anonimizado_1.xlsx")
uva.a.1819 <- read_xlsx("Data/uva-acceso-2018-19-anonimizado_1.xlsx")
uva.a.1920 <- read_xlsx("Data/uva-acceso-2019-20-anonimizado_1.xlsx")
uva.a.2021 <- read_xlsx("Data/uva-acceso-2020-21-anonimizado_1.xlsx")
uva.a.2122 <- read_xlsx("Data/uva-acceso-2021-22-anonimizado_0.xlsx")
uva.a.2223 <- read_xlsx("Data/uva-acceso-2022-23-anonimizado.xlsx")
uva.a <- rbind(uva.a.1718,uva.a.1819,uva.a.1920,uva.a.2021,uva.a.2122,uva.a.2223)
rm(uva.a.1718,uva.a.1819,uva.a.1920,uva.a.2021,uva.a.2122,uva.a.2223)
uva.a$uni.madre <- (uva.a$des_nivel_estudios_madre=="Estudios Superiores")
uva.a$uni.padre <- (uva.a$des_nivel_estudios_padre=="Estudios Superiores")
uva.a$nota_admision[uva.a$cod_titulacion==2502312] <- 5
hh <- uva.a %>% group_by(cod_titulacion) %>%
  summarize(n.acceso=n(),
            nota.mediana=median(nota_admision,na.rm=TRUE),
            nota.min=min(nota_admision,na.rm=TRUE),
            uni.madre=sum(uni.madre,na.rm=TRUE)/n.acceso,
            uni.padre=sum(uni.padre,na.rm=TRUE)/n.acceso)
bd.uva <- merge(bd.uva,hh,by="cod_titulacion",all.x = TRUE)

# UC3M
# =====
uc3m.a.1718 <- read_xlsx("Data/uc3m-acceso-2017-18-anonimizado_0.xlsx")

```

```

uc3m.a.1819 <- read_xlsx("Data/uc3m-acceso-2018-19-anonimizado_0.xlsx")
uc3m.a.1920 <- read_xlsx("Data/uc3m-acceso-2019-20-anonimizado_0.xlsx")
uc3m.a.2021 <- read_xlsx("Data/uc3m-acceso-2020-21-anonimizado_0.xlsx")
uc3m.a.2122 <- read_xlsx("Data/uc3m-acceso-2021-22-anonimizado.xlsx")
# uc3m.a.2223 <- read_csv("Data/uc3m-acceso-2022-23-anonimizado.xlsx")
uc3m.a <- rbind(uc3m.a.1718,uc3m.a.1819,uc3m.a.1920,uc3m.a.2021,uc3m.a.2122) #,uc3m.a.2223)
rm(uc3m.a.1718,uc3m.a.1819,uc3m.a.1920,uc3m.a.2021,uc3m.a.2122)
uc3m.a$uni.madre <- (uc3m.a$des_nivel_estudios_madre=="Estudios Superiores")
uc3m.a$uni.padre <- (uc3m.a$des_nivel_estudios_padre=="Estudios Superiores")
uc3m.a$nota_admision[uc3m.a$des_centro=="Centro Universitario Guardia Civil"] <- 5
hh <- uc3m.a %>% group_by(cod_titulacion) %>%
  summarize(n.acceso=n(),
            nota.mediana=median(nota_admision,na.rm=TRUE),
            nota.min=min(nota_admision,na.rm=TRUE),
            uni.madre=sum(uni.madre,na.rm=TRUE)/n.acceso,
            uni.padre=sum(uni.padre,na.rm=TRUE)/n.acceso)
bd.uc3m <- merge(bd.uc3m,hh,by="cod_titulacion",all.x = TRUE)

# URJC
# =====
urjc.a.1718 <- read_xlsx("Data/urjc-acceso-2017-18-anonimizado_0.xlsx")
urjc.a.1819 <- read_xlsx("Data/urjc-acceso-2018-19-anonimizado_0.xlsx")
urjc.a.1920 <- read_xlsx("Data/urjc-acceso-2019-20-anonimizado_0.xlsx")
urjc.a.2021 <- read_xlsx("Data/urjc-acceso-2020-21-anonimizado_0.xlsx")
urjc.a.2122 <- read_xlsx("Data/urjc-acceso-2021-22-anonimizado_0.xlsx")
urjc.a.2223 <- read_xlsx("Data/urjc-acceso-2022-23-anonimizado.xlsx")
urjc.a <- rbind(urjc.a.1718,urjc.a.1819,urjc.a.1920,urjc.a.2021,urjc.a.2122,urjc.a.2223)
rm(urjc.a.1718,urjc.a.1819,urjc.a.1920,urjc.a.2021,urjc.a.2122,urjc.a.2223)
urjc.a$uni.madre <- (urjc.a$des_nivel_estudios_madre=="Estudios Superiores")
urjc.a$uni.padre <- (urjc.a$des_nivel_estudios_padre=="Estudios Superiores")
hh <- urjc.a %>% group_by(cod_titulacion) %>%
  summarize(n.acceso=n(),
            nota.mediana=median(nota_admision,na.rm=TRUE),
            nota.min=min(nota_admision,na.rm=TRUE),
            uni.madre=sum(uni.madre,na.rm=TRUE)/n.acceso,
            uni.padre=sum(uni.padre,na.rm=TRUE)/n.acceso)
bd.urjc <- merge(bd.urjc,hh,by="cod_titulacion",all.x = TRUE)

```

Estructura del PDI

```

uam.pdi <- read_xlsx("Data/uam-personal-pdi-2022-anonimizado.xlsx")
uc3m.pdi <- read_xlsx("Data/uc3m-personal-pdi-2022-anonimizado.xlsx")
ucm.pdi <- read_xlsx("Data/ucm-personal-pdi-2022-anonimizado.xlsx")

```



```

uva.pdi <- read_xlsx("Data/uva-personal-pdi-2022-anonimizado.xlsx")
urjc.pdi <- read_xlsx("Data/urjc-personal-pdi-2022-anonimizado.xlsx")
pdi <- rbind(uam.pdi,uc3m.pdi,ucm.pdi,uva.pdi,urjc.pdi)[,c(2,3,23,26,27)]

xx <- table(pdi$des_categoria_cuerpo_escalas,pdi$cod_universidad)
xx <- matrix(xx, ncol = ncol(xx), dimnames = dimnames(xx))
xx <- (t(xx)/colSums(xx))
xx <- data.frame(cod_universidad=row.names(xx),xx)

```

Unión bases de las 5 universidades

Unión bases de datos de las 5 universidades

```

bd <- rbind(bd.ucm,bd.uam,bd.uva,bd.uc3m,bd.urjc)
# join PDI
bd <- merge(bd,xx, by="cod_universidad",all.x = TRUE)

hh <- bd %>% group_by(des_titulacion,cod_universidad) %>%
  summarize(n=n(),
            cre=mean(num_total_creditos_mat_curso,na.rm=T),
            max=max(num_total_creditos_mat_curso,na.rm=T),
            min=min(num_total_creditos_mat_curso,na.rm=T))

bd <- bd[(bd$num_total_creditos_mat_curso < 91) &
        (bd$num_total_creditos_mat_curso >= 30),]
bd$des_titulacion <- as.factor(bd$des_titulacion)
bd$cod_titulacion <- as.factor(bd$cod_titulacion)
bd <- bd[!is.na(bd$nota.mediana),]

```

Tasa de Rendimiento

Calculo de la tasa de rendimiento y clasificación de estudiantes (0/1).

```

# Tasa rendimiento
bd$tr <- bd$num_total_creditos_sup_curso/bd$num_total_creditos_mat_curso
# # Tasa éxito
# bd$te <- bd$num_total_creditos_sup_curso/bd$num_creditos_pres_curso
# # Tasa evaluación
# bd$tev <- bd$num_creditos_pres_curso/bd$num_total_creditos_mat_curso

bd$abandono <- as.factor(bd$tr <= 18/60)
levels(bd$abandono) <- c("NO", "SI")
table(bd$abandono)

```

```
##
```



```
##      NO      SI
## 184988 18953

levels(bd$des_rama) <- c("AH","Ci","CS","CSJ","IA")
```

Algoritmo MARS

```
formula <- abandono ~ des_genero + doble + des_rama + adscrito + cod_universidad +
  n.titula + n.acceso + edad.media.tit + muni.local + prov.local + nota.mediana +
  nota.min + uni.padre + uni.madre + Profesor.Asociado + Profesor.Ayudante.Doctor
mymars <- earth(formula, data = bd, minspan=500, endspan =
  500,degree=2,glm=list(family=binomial))
summary(mymars)
```

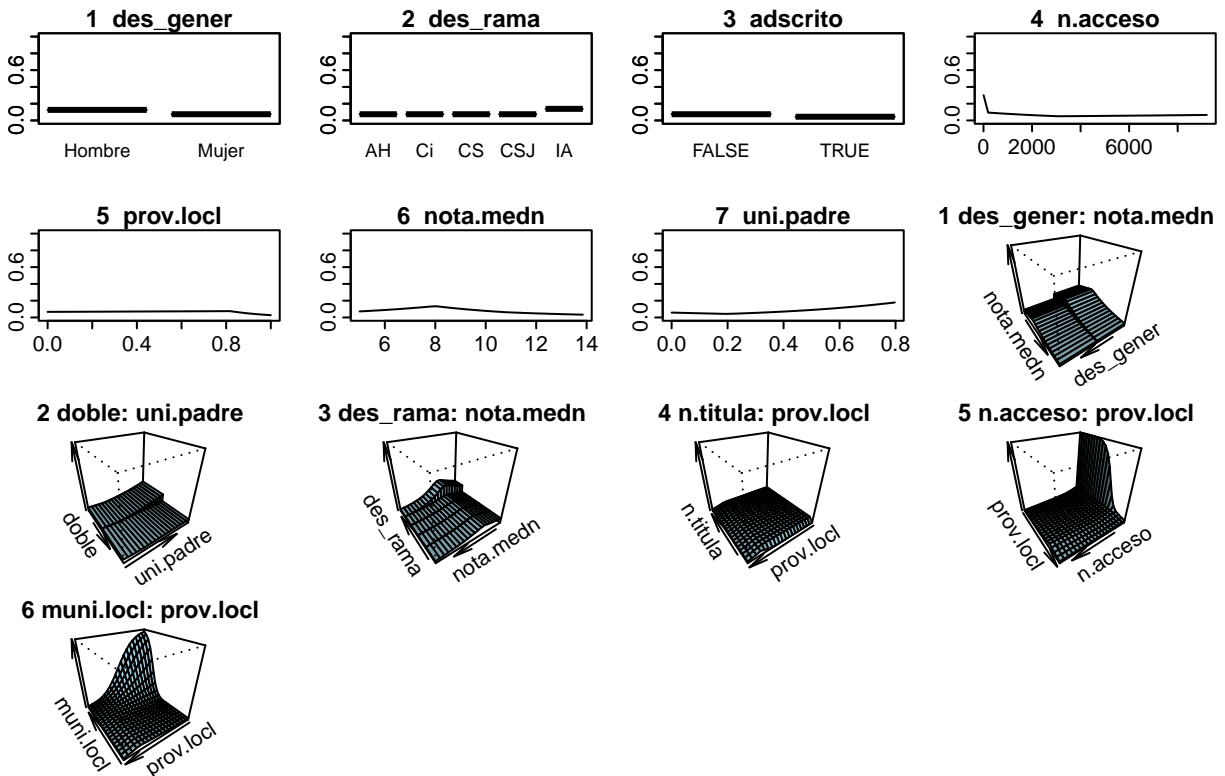
```
## Call: earth(formula=formula, data=bd, glm=list(family=binomial), degree=2,
##           minspan=500, endspan=500)
##
## GLM coefficients
##
##                                     SI
## (Intercept)                        -2.0980746
## des_generoMujer                    -0.5443135
## des_ramaIA                         1.0642085
## adscritoTRUE                      -0.5557541
## h(3090-n.acceso)                   0.0002066
## h(n.acceso-3090)                  0.0000745
## h(0.818589-prov.local)             -1.1720200
## h(prov.local-0.818589)             -5.8662504
## h(8-nota.mediana)                 -0.1735601
## h(nota.mediana-8)                 -0.3656122
## h(0.198732-uni.padre)              1.6966739
## h(uni.padre-0.198732)              2.6620955
## des_generoMujer * h(nota.mediana-10.862) 0.1496124
## des_generoMujer * h(10.862-nota.mediana) -0.0607843
## dobleTRUE * h(uni.padre-0.198732)      -2.1478878
## des_ramaIA * h(nota.mediana-8.764)     -0.2628708
## des_ramaIA * h(8.764-nota.mediana)      0.0173064
## h(90-n.titula) * h(0.818589-prov.local) -0.1880761
## h(83-n.acceso) * h(0.818589-prov.local) 0.2036405
## h(n.acceso-83) * h(0.818589-prov.local) -0.0003117
## h(muni.local-0.323782) * h(0.818589-prov.local) 27.5316395
##
## GLM (family binomial, link logit):
## nulldev      df      dev      df  devratio    AIC iters converged
##   126147 203940   114570 203920    0.0918 114600     6           1
##
```

```
## Earth selected 21 of 23 terms, and 10 of 22 predictors
## Termination condition: RSq changed by less than 0.001 at 23 terms
## Importance: des_ramaIA, nota.mediana, des_generoMujer, muni.local, ...
## Number of terms at each degree of interaction: 1 11 9
## Earth GCV 0.07850917    RSS 16003.23    GRSq 0.06866982    RSq 0.06912643
```

```
plotmo(mymars)
```

```
## plotmo grid:    des_genero doble des_rama adscrito cod_universidad n.titula
##                Mujer FALSE      CSJ      FALSE          010      1036
## n.acceso edad.media.tit muni.local prov.local nota.mediana nota.min uni.padre
##      1281      22.24598  0.3739353  0.7316176          10.179      5.007  0.4233577
## uni.madre Profesor.Asociado Profesor.Ayudante.Doctor
##      0.471831          0.220434          0.1129216
```

```
SI earth(formula, data=bd, glm=list(family=binomial), degree=2, minspan=500,...
```



Referencias

- Constante-Amores, Alexander, Eva Florenciano Martinez, Enrique Navarro Asencio, and Maria Fernandez-Mellizo. 2021. "Factores Asociados Al Abandono Universitario." *Educacion XX1* 24 (1): 17–44.
- Kehm, Barbara M, Malene Rode Larsen, and Hanna Bjornoy Sommersel. 2019. "Student Dropout from Universities in Europe: A Review of Empirical Literature." *Hungarian*

Educational Research Journal 9 (2): 147–64.

Llave Montiel, Miguel Angel de la, and Fernando Lopez. 2020. “Spatial Models for Online Retail Churn: Evidence from an Online Grocery Delivery Service in Madrid.” *Papers in Regional Science* 99 (6): 1643–66.

Ortiz-Lozano, Jose Maria, Antonio Rua-Vieites, Paloma Bilbao-Calabuig, and Marti Casadesus-Fa. 2020. “University Student Retention: Best Time and Data to Identify Undergraduate Students at Risk of Dropout.” *Innovations in Education and Teaching International*.