

# Modelado de variables demograficas mediante algoritmos de aprendizaje supervisado

## Introducción

Cuando se trabaja con modelos de micro-simulación en demografía, como puede ser el modelado de las tasas específicas de fecundidad por edad (ASFR), uno de los factores usualmente contemplado es la edad en que las mujeres comienzan a tener riesgo a concebir. En ese sentido, modelar la edad en que una cohorte de mujeres experimenta su primer nacimiento o cuando comienzan a tener relaciones sexuales es un factor que puede ser de utilidad para este tipo de simulación.

En este caso se explora el uso de un modelo de regresión para predecir la edad al primer hijo de las mujeres y otro de clasificación para predecir si una mujer comienza a tener relaciones sexuales antes o después de cierta edad. Se utilizarán principalmente modelos basados en árboles de decisión, como CART, Random Forest, XGBoost y Bagging.

En este caso se trabaja con datos de las encuestas DHS, realizada principalmente en países en vías de desarrollo, donde se tomarán los datos de las encuestas VI y VII, realizadas entre los años 2010-2020, se utilizarán los datos de varios países de distintas regiones para visualizar el efecto de las variables de interés en presencia de contextos socioculturales y económicos muy distintos.

## Fuente de los datos

Como se menciona en la introducción los datos provienen de las encuestas DHS, el Programa de Encuestas Demográficas y de Salud (DHS) es responsable de recopilar y difundir datos precisos y representativos a nivel nacional sobre salud y población en países en desarrollo. El proyecto es financiado por la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID).

La idea detrás es ayudar a los distintos países sobre todo del continente africano a dar asistencia técnica para la implementación de encuestas demográficas y de salud, donde se recolecta

información sobre la fecundidad, planificación familiar, salud maternal e infantil y otros temas de salud como la malaria o el VIH que son de interés en gran parte de los países que contempla la encuesta.

## Presentación de los datos y análisis exploratorio

Los datos utilizados en esta ocasión corresponden a 12 países de distintas regiones del mundo:

- America Latina y el Caribe: Peru, Haiti, Guatemala y República Dominicana.
- Europa del este: Albania y Armenia.
- Norte y Sur de Africa: Etiopia, Ghana, Mali, Nigeria, Sudáfrica y Angola.

En este caso se tiene un total de casi 200 mil observaciones que corresponden a mujeres de entre 15 y 49 años, donde se tienen variables de aspectos economicos, sociales y de salud reproductiva. En este caso se trabaja con 17 variables en total que se detallan acontinuacion:

Table 1: Descripción de las variables utilizadas en el trabajo

Codigo DHS	Nombre	Descripción	Tipo	Recorrido
v000	pais	País donde se realizó la encuesta	Categórica	12 países
v005	ponderador	Peso de la mujer en la encuesta, número entero de 8 dígitos, donde 6 dígitos corresponden a decimales	Numérica Entera	
v007	anio	Año en que se realizó la encuesta	Numérica Entera	2008-2018
v012	edad	Edad de la mujer encuestada	Numérica Entera	15-64
v102	urbano	Indica si la mujer es de área urbana o rural	Categórica	1:Urbano, 2:Rural
v119	electricidad	Indica si el hogar tiene acceso a electricidad	Categórica	0: No, 1: Si
v121	tv	El hogar dispone de una televisión	Categórica	0: No, 1: Si
v133	anio_educ	Total años de educación de la mujer, se calcula a partir del nivel máximo de educación alcanzado, es un valor comparable entre países.	Numérica Entera	0-25
v136	tot_fam	Total de personas que viven en el hogar	Numérica Entera	1-39
v157	leer	Frecuencia con que la mujer lee un periódico, revista o libro	Categórica	0: Nada, 1: Ocasional, 2: Todos los días
v191	ind_riqueza	Indicador de riqueza del hogar estandarizado, se calcula a partir de distintas medidas como materiales de construcción de la casa y servicios disponibles, su calculo puede variar entre países.	Numérica Entera	-925630-5232460
v212	edad_phijo	Edad de la mujer al primer hijo	Numérica Entera	3-47
v301	con_antic	Conocimiento de algún método anticonceptivo	Categórica	0: Ninguno, 1:Folklórico, 2:Tradicional, 3:Moderno
v525	edad_psexo	Edad de la mujer al primer acto sexual	Numérica Entera	0-49
v714	trabajo	Indica si la mujer actualmente trabaja	Categórica	0: No, 1: Si
v836	n_parejas	Cantidad de parejas sexuales que ha tenido la mujer	Numérica Entera	1-95