

1 Objetivos

En esta práctica utilizaremos los datos abiertos de PDI de la UAM correspondiente al año 2019, disponible en [este link](#) con el objetivo de identificar al profesor Ortigosa. Para ello se nos ha proporcionado su explícito consentimiento. En particular, inferiremos el valor de los aquellos campos que no empiezan por *cod_*, *lon_* o *lat_*.

Para poder estudiar este *dataset* hemos de conocer en detalle el proceso de anonimización seguido para crearlo. Podemos encontrarlo en [este link](#).

2 Información utilizada

Se nos ha proporcionado la siguiente información sobre el profesor Ortigosa:

- El profesor Ortigosa estaba en 2019 en la UAM. Esto es, tenemos asegurado que aparece en el fichero.
- El profesor Ortigosa era de género masculino en 2019.
- El profesor Ortigosa pertenecía al Departamento de Ingeniería Informática en 2019.

Adicionalmente, utilizaremos las siguientes piezas de información encontradas buscando sobre el profesor Ortigosa en internet.

- En [la Biblioteca-UAM](#) encontramos la tesis doctoral del profesor Ortigosa. En particular nos interesa su año de publicación, **2000**, y la universidad en la que se doctoró, la **Universidad Autónoma de Madrid**.
- En [portal de miembros de la Universidad Autónoma de Madrid](#) podemos encontrar el cargo del profesor Ortigosa: **Ingeniero Informático, Profesor Titular**.
- En el [LinkedIn del profesor Ortigosa](#) podemos encontrar sus numerosos cargos actuales además del cargo de profesor titular.
- En la página 20 de los [Estatutos de la UAM](#) encontramos la definición de **cargo unipersonal**.
- En la [Oficina de prácticas de la EPS](#) descubrimos que **el profesor Ortigosa no es director, subdirector ni secretario del Departamento de Ingeniería Informática** al que pertenece. Si que es considerado **Miembro Nato del Consejo de Departamento**.
- En el [perfil de investigador del profesor Ortigosa](#) encontramos información relacionada con su carrera académica en la UAM. Nos es particularmente relevante el Área de Conocimiento, **Lenguajes Y Sistemas Informáticos**, así como su número de quinquenios y sexsenios: 3 y 4 respectivamente. Además, podemos inferir el año de incorporación al cuerpo docente a partir de la primeras asignaturas que impartió en la UAM en el año 2008.
- Hemos consultado [Documat](#) y [ResearchGate](#) para investigar si el profesor Ortigosa había publicado una segunda tesis, pero no hemos encontrado evidencias de ello.
- En [la página de equipo GHIA](#) encontramos al profesor Ortigosa como miembro del equipo, no quedando claro si es investigador principal de algún proyecto.

Asumiremos que todas estas fuentes de información son precisas y están actualizadas en la fecha de consulta, 27 de Diciembre de 2021.

3 Proceso seguido

En primer lugar hemos de identificar las variables pivote en el proceso de anonimización. Estudiando el proceso de anonimización sabemos que estas son *cod_unidad_responsable/des_unidad_responsable* y *cod_genero/des_genero*. Sabemos por un lado que el género del profesor Ortigosa es masculino. Por otro lado, la variable *des_unidad_responsable* hace referencia al departamento, también conocido: Departamento de Ingeniería Informática.

Conociendo el proceso de anonimización sabemos que los valores de cada variable para el profesor Ortigosa han de tener asociadas las variables pivotes. Es decir, podemos restringirnos a aquellas filas del *dataset* con estos valores en los campos *des_genero* y *des_unidad_responsable*.

Obtenemos que únicamente 56 filas de las originales 2577 cumple este criterio. Estudiaremos los valores que puede tomar cada campo del *dataset* para el profesor Ortigosa por bloques de coherencia, ya que son aquellas variables que permanecen relacionadas tras las permutación.

3.1 Bloque de coherencia 1

El primer bloque de coherencia incluye las variables referentes a la universidad y el año en el que este campo ha sido añadido. En este caso la estimación es sencilla pues las 56 filas toman los mismos valores:

Campo	Valor estimado	Nivel de Certeza (%)
des_universidad	Universidad Autónoma de Madrid	100%
anio	2019	100%

3.2 Bloque de coherencia 2

Las variables del segundo bloque de coherencia son *des_pais_nacionalidad*, *des_continente_nacionalidad* y *des_agregacion_paises_nacionalidad*. Estudiamos la frecuencia con la que aparece cada valor en nuestro *dataset* reducido con las variables pivote y asignaremos al profesor Ortigosa aquel valor con mayor frecuencia. El nivel de certeza será la frecuencia de aparición de dicho valor. Seguiremos este mismo proceso para el resto de bloques de coherencia. Obtenemos los siguientes resultados:

Campo	Valor estimado	Nivel de Certeza (%)
des_pais_nacionalidad	Universidad Autónoma de Madrid	98.21%
des_continente_nacionalidad	Europa meridional	100%
des_agregacion_paises_nacionalidad	Europa	100%

3.3 Bloque de coherencia 3

Repetimos el proceso para las variables del tercer bloque de coherencia:

Campo	Valor estimado	Nivel de Certeza (%)
des_comunidad_residencia	Madrid	100%
des_provincia_residencia	Madrid	100%
des_municipio_residencia	MADRID	51.78%

3.4 Bloque de coherencia 4

El cuarto bloque de coherencia únicamente incluye la fecha de nacimiento. Para este campo los valores tomados por nuestro *dataset* son muy dispares. De cara a obtener una mejor predicción hemos encontrado el año de publicación de su tesis doctoral: 2000. Sabiendo que antiguamente las licenciaturas se terminaban con una media de 23 años, y añadiendo 3 adicionales para la tesis obtenemos una edad mínima para la publicación de la tesis de 26 años. Siendo conservadores diremos que el profesor Ortigosa nació antes del año 1975.

Aplicando este nuevo filtro a los datos obtenemos una estimación más precisa. En particular obtenemos el mismo valor con mayor frecuencia, pero dicha frecuencia es más alta, luego nuestro nivel de certeza es mayor. Estos son los resultados obtenidos:

Campo	Valor estimado	Nivel de Certeza (%)
anio_nacimiento	1967	13.88%

3.5 Bloque de coherencia 5

El quinto bloque de coherencia incluye variables asociadas al cargo del profesor Ortigosa en la UAM. En particular, conocemos el cargo del profesor Ortigosa (Profesor Titular), así que podemos restringirnos a las filas que toman dicho valor. Obtenemos así los siguientes resultados:

Campo	Valor estimado	Nivel de Certeza (%)
des_tipo_personal	Funcionario Interino	100%
des_categoria_cuerpo_escal	Profesor Titular de Universidad	100%
des_tipo_contrato	Personal Funcionario\Personal eventual	100%
des_dedicacion	Dedicación a Tiempo Completo	100%
num_horas_semanales_tiempo_parcial	NaN	100%
des_situacion_administrativa	Servicio Activo	92.85%

Hemos de aclarar que el número de horas semanales a tiempo parcial aparece como *NaN* pues si el contrato es de tiempo completo, este campo toma dicho valor. Aclaremos también que la investigación realizada ha revelado que el profesor Ortigosa posee numerosos cargos adicionales además de su puesto de profesor titular en la universidad. Sin embargo, escapa a nuestro conocimiento si estos cargos interfieren de ninguna forma con su ocupación en la universidad y si influye en estos campos. Es por ello que no hemos utilizado esta información para nuestra inferencia.

3.6 Bloque de coherencia 6

El sexto bloque de coherencia contiene una única variable booleana indicando si el individuo ostenta un cargo unipersonal remunerado en la universidad. Aunque sabemos que el profesor Ortigosa es profesor titular, ¿es esto un cargo unipersonal? Revisando los estatutos de la universidad encontramos la definición de cargo unipersonal: *Rector; Vicerrectores, Secretario General, Gerente; Decanos, Vicedecanos y Secretarios de Facultad; Directores, Subdirectores y Secretarios de Escuela; Directores y Secretarios de Departamento; Directores de Instituto Universitario de Investigación; y Administradores Gerentes de Centro.*

Con nuestro conocimiento del problema sabemos que el profesor Ortigosa no es director, subdirector ni secretario del departamento al que pertenece. Si es miembro nato del consejo de departamento, pero este no se considera un cargo unipersonal. Por lo tanto, no podemos asegurar que el profesor Ortigosa ostente este tipo de cargo. Nos remitimos entonces a la inferencia realizada hasta ahora a partir de la frecuencia de nuestro dataset, obteniendo el siguiente resultado:

Campo	Valor estimado	Nivel de Certeza (%)
ind_cargo_remunerado	N	82.14%

Donde el valor *N* significa que no ostenta dicho cargo.

3.7 Bloque de coherencia 7

El séptimo bloque de coherencia incluye variables relaciones con el título de doctor. A partir de las tesis encontrada sabemos que ésta se presentó en el año 2000 a través de la Universidad Autónoma de Madrid. Filtrando en nuestro *dataset* utilizando estos valores obtenemos una buena predicción en todas las variables de este bloque:

Campo	Valor estimado	Nivel de Certeza (%)
des_titulo_doctorado	Uno	100%
des_pais_doctorado	España	100%
des_continente_doctorado	Europa	100%
des_agregacion_paises_doctorado	Europa meridional	100%
des_universidad_doctorado	Universidad Autónoma de Madrid	100%
anio_lectura_tesis	2000.0	100%
anio_expedicion_titulo_doctor	2000.0	100%
des_mencion_europea	No	100%

3.8 Bloque de coherencia 8

El octavo bloque de coherencia incluye variables relacionadas con el departamento del profesor Ortigosa. En particular, sabemos tanto el departamento como el área de conocimiento del profesor Ortigosa a partir de nuestra búsqueda de información previa:

Campo	Valor estimado	Nivel de Certeza (%)
des_tipo_unidad_responsable	Departamento	100%
des_area_conocimiento	Lenguajes y Sistemas Informáticos	100%

3.9 Bloque de coherencia 9

El noveno bloque de coherencia contiene variables relacionadas con el tiempo que lleva trabajando el personal de la Universidad en la misma. En este caso filtraremos nuestra información conociendo número quinquenios (3) y sexsenios (4) que ha trabajado el profesor Ortigosa en la UAM. Además, sabiendo que impartió su primeras asignaturas en el año 2008, asumiremos que este es el año de incorporación del profesor Ortigosa al sistema como docente.

Sin embargo, para los campos filtrados con 3 quinquenios y 4 sexsenios encontramos valores *NaN* en el campo *anio_incorporacion_ap* para todas las filas, por lo que suprimimos el filtro asociado al año de incorporación a la docencia.

Fijando estos valores para los respectivos campos, inferimos el resto de campos del bloque de coherencia:

Campo	Valor estimado	Nivel de Certeza (%)
anio_incorporacion_ap	NaN	100%
anio_incorpora_cuerpo_docente	NaN	100%
num_trienios	5/6/7	33.33%
num_quinquenios	3	100%
num_sexsenios	4	100%

Donde hemos obtenido 3 valores distintos para el campo *num_trienios* con la misma probabilidad.

3.10 Bloque de coherencia 10

El décimo bloque de coherencia contiene una única variable que notifica del número de tesis doctorales que el individuo ha realizado. En nuestro caso, sabemos que el profesor Ortigosa ha publicado al menos una tesis, y en el *dataset* en estudio no hay casos con más de dos, así que nos preguntamos si el profesor Ortigosa ha escrito una segunda tesis.

Sin embargo, no hemos encontrado constancia en la investigación realizada de que así sea. Puede darse el caso que dicha tesis no haya sido publicada públicamente o no este disponible en los portales consultados, así que no asumiremos que el profesor Ortigosa ha realizado únicamente una tesis.

Aún así, podemos filtrar nuestro datos utilizando que ha escrito *al menos una* tesis doctoral. Obtenemos la siguiente estimación:

Campo	Valor estimado	Nivel de Certeza (%)
num_tesis	1.0	71.42%

3.11 Bloque de coherencia 11

El último bloque de coherencia incluye una única variable booleana señalando si el individuo ha sido el Investigador Principal (IP) de algún proyecto o contrato durante el año. En nuestra investigación hemos encontrado el equipo del profesor Ortigosa, el GHIA. Sin embargo, no queda claro si el profesor Ortigosa ha sido IP del algún proyecto o contrato este año, por lo que nos remitimos al proceso de inferencia realizado hasta ahora:

Campo	Valor estimado	Nivel de Certeza (%)
num_tesis	N	78.57%