

IT Academy - Big Data (online)

Memoria del Proyecto Big Data:

Llamadas de la linea de atención contra la violencia de género.

Juan Javier Hidalgo Gómez

Octubre 2023

Introducción:

1. Obtención de datos.
2. Descripción de los datos.
3. Visualización de datos.
4. Modelo de Agrupación.
5. Conclusiones.

Introducción.

En este trabajo fin de curso se pretende poner en práctica los conocimientos y técnicas aprendidas en el transcurso del curso Big Data de IT Academy mediante su aplicación práctica en un conjunto de datos públicos.

Me planteo dos objetivos principales: por una parte consolidar el uso de las técnicas y la interpretación de los resultados y por otro desarrollar las habilidades necesarias que permiten la aplicación de estas herramientas en el mundo real y en casos reales de negocio.

Entiendo este segundo objetivo como la constatación efectiva del aprovechamiento de los conocimientos adquiridos. Más allá la amplitud de conocimientos y destreza en el uso de técnicas y métodos en el ámbito del Big Data, éstos no son más que el medio para permitirnos conocer mejor el escenario en el que desarrollamos nuestra actividad y ofrecer una toma de decisiones adecuada a los objetivos marcados en el negocio.

Dividiremos este documento en varias partes. En el apartado de obtención de datos comentaremos cuál ha sido el origen de datos, una breve descripción sobre el conjunto de datos escogido y las preparaciones básicas realizadas para favorecer su posterior tratamiento y análisis que haremos.

La parte de preparación de datos realizará un recorrido exhaustivo por los diferentes atributos analizando su contenido y distribución y haciendo, en su caso, adecuaciones para evitar influencias no deseadas en los análisis posteriores.

Los modelos de agrupación, como modelos de aprendizaje no supervisado, nos ofrecerán la posibilidad de agrupar las instancias de nuestros datos en función de sus características.

1. Obtención de datos

En esta parte cargaremos el conjunto de datos que queremos analizar y realizaremos un primer análisis y preparación de datos consistente en el tratamiento de nulos y duplicados, y en la adecuación de las variables o atributos. En este sentido se han adaptado los nombres de las columnas para facilitar su análisis

y se han identificado las variables categóricas en las que se han mantenido columnas descriptivas y numéricas. Esta primera parte empieza a conocerse el significado de los datos que se profundizará en la siguiente parte con la descripción de los atributos y visualizaciones realizadas.

Origen de dades: <https://datos.gob.es/es/catalogo/a09002970-linea-de-atencion-contra-la-violencia-machista-900-900-120>

Conjunto de datos del Ministerio de asuntos económicos y transformación digital.

La finalidad de este dataset:

- Análisis de las Llamadas de la línea de atención contra la violencia de género.

2. Descripción de los datos

El Dataset es un formulario/cuestionario de preguntas que se realizan a la persona que realiza la llamada, cuya finalidad es buscar un perfil de la misma. Por lo que nos encontramos una serie de atributos categóricos principalmente, exepctuando los que identifican la fecha de la llamada.

Información de los campos:

<https://analisi.transparenciacatalunya.cat/Societat-benestar/L-nia-d-atenci-contra-la-viol-ncia-masclista-900-9/q2sg-894k>

Data: Data de la trucada. '%d/%m/%Y'.

Any: Any que s'ha realitzat la trucada.

Durada segons: Intèrval de la durada de la trucada.

Día setmana: Dia de la setmana que s'ha realitzat la trucada.

Mes: Mes que s'ha realitzat la trucada.

Hora: Hora de la trucada.

Llengua: Idioma en què s'ha efectuat la trucada.

Comarca: Comarca d'on procedeix la trucada.

Àmbit: territorial: **Àmbit territorial d'on procedeix la trucada.** Província: **Província d'on procedeix la trucada.** Província agrupada: **Província des d'on truca la persona agrupada en categories.** Sexe: **Sexe de la persona que truca o sobre qui es fa la consulta.** Edat: **Edat de la persona que truca o sobre qui es fa la consulta agrupada en categories.** Estat Civil: **Detall de la persona professional que truca en relació a la xarxa de recursos públics.** Persona que truca: **Persona que truca i relació amb la persona usuària.** Detall familiar o persona proopera: **Més detalls de la persona que truca, si s'escau.** Detall professional: **Detall de la persona professional que truca.** Situació familiar: **Situació familiar de la persona que truca o sobre qui es fa la consulta.** Situació fills/filles: **Situació en relació a les filles/s de la persona que truca o sobre qui es fa la consulta.** Nombre fills/filles: **Nombre de fills de la persona que truca o sobre qui es fa la consulta.** Situació fills/filles A: **Situació en relació als fills de la persona que truca o sobre qui es fa la consulta** Fa referència a una segona situació en relació als fills. Nombre fills/filles A: **Nombre de filles/s de la persona que truca o sobre qui es fa la consulta agrupada en categories.** Fa referència a una segona situació en relació als fills. Pateixen maltractament: **Situació de maltractament de les filles/s dins del nucli familiar de la persona que truca o sobre qui es fa la consulta.** Nacionalitat: **Nacionalitat de la persona que truca o sobre qui es fa la consulta.** Situació legal: **Situació econòmica de la persona que truca o sobre qui es fa la consulta.** Temps residència: **Anys de residència de la persona estrangera que truca o sobre qui es fa la consulta agrupada.** Formació: **Formació de la persona que truca o sobre qui es fa la consulta.** Situació Laboral: **Situació laboral de la persona que truca o de qui es fa la**

consulta. Detall situació laboral: **Detall de la situació laboral de la persona que truca o sobre qui es fa la consulta.** Ús previ servei: **Ús previ del servei de la persona que truca o sobre qui es fa la consulta.** Motiu Trucada: **Motiu de la trucada.** Tipus assessorament: **Tipus d'assessorament de la trucada en el cas que no es tracti de violència masclista.** Àmbit violència masclista: **Més informació sobre l'àmbit de la violència masclista.** Detall àmbit: **Més informació sobre l'àmbit del motiu de la trucada.** Violència física: **Detalla si es pateix violència física.** Violència psicològica: **Detalla si es pateix violència psicològica.** Violència sexual: **Detalla si es pateix violència sexual.** Violència econòmica: **Detalla si es pateix violència econòmica.** Conviu agressor: **Detalla si la persona que truca o de qui es fa la consulta conviu amb la persona agressora.** Relació agressor: **Relació de la persona amb l'agressor.** Relació agressor àmbit laboral: **Relació de la persona amb l'agressor en l'àmbit laboral.** Temps de relació: **Temps que fa que dura la relació.** Nacionalitat agressor: **Nacionalitat de la persona agressora.** Derivació: **Detalla si s'ha derivat la trucada a algun altre servei.** Xarxa: **Detalla el recurs específic a on s'ha derivat la trucada.** Derivació_A: **Detalla si s'ha derivat la trucada a algun altre servei.** Xarxa_A: **Detalla el recurs específic a on s'ha derivat la trucada.** Derivació_B: **Detalla si s'ha derivat la trucada a algun altre servei.** Xarxa_B: **Detalla el recurs específic a on s'ha derivat la trucada.** Derivació_C: **Detalla si s'ha derivat la trucada a algun altre servei.** Xarxa_C: **Detalla el recurs específic a on s'ha derivat la trucada.** Traducció: **Detalla si ha calgut traducció per atendre la trucada.** Idioma Traducció: **Idioma de la traducció.** Reclamacions/Queixes: **Detalla si hi ha hagut reclamacions o queixes.** Emergència: **Detalla si es tracta d'una emergència.** Connexió: **Tipus de connexió:** 1-Watsapp, 2-Trucades i 3-Correu electrònic.

3. Visualización de datos.

En este apartado se recorrerán cada uno de los atributos del dataset. La finalidad es identificar el rango de valores y demás descriptores estadísticos, así como identificar la existencia de outliers y decidir que hacer con ellos en función de la influencia que amamos pueden tener en cuanto a distorsiones en la interpretación de los resultados. Por esta tarea nos apoyaremos en visualizaciones simples como boxplot e histogramas básicamente.

4. Modelo de Agrupación.

El objetivo de nuestro proyecto es realizar una agrupación/categorización de las llamadas por el tipo de cuestionario que se ha rellenado al realizar la llamada.

En este aparte dedicado al aprendizaje no supervisado desarrollaremos los métodos K-means para hacer una agrupación de las instancias en clusters de forma que las que se encuentren en el mismo cluster estén entre ellas más cerca de lo que estarían con instancias de otros clusters. Utilizaremos el análisis principal de componentes para vernos su bondad y cómo reducir los atributos a considerar a la hora de realizar una agrupación. Además de la definición, entrenamiento y prueba del modelo, veremos diferentes formas de calcular el valor óptimo de clusters como puede ser el método del codo, el dendograma y el método de Silhouette.