

**Práctica 2 Inteligencia de Negocio**

Rafael Muñoz Olmedo

Curso académico 2020|2021

Índice

[1. Introducción 2](#_Toc61300924)

[2. Casos de estudio 3](#_Toc61300925)

[2.1. Caso 1: colisiones de vehículos con más de un muerto 3](#_Toc61300926)

[2.2. Caso de estudio 2: accidentes sin víctimas mortales 3](#_Toc61300927)

[2.3. Caso de estudio 3: accidentes con 10 o más víctimas 4](#_Toc61300928)

[2.4. Caso de estudio 4: accidentes con mal tiempo 4](#_Toc61300929)

[2.5. Caso de estudio 5: colisión de vehículos 4](#_Toc61300930)

[2.6. Análisis de los resultados 5](#_Toc61300931)

[2.7. KNN con nuevos ajustes 5](#_Toc61300932)

# 

# Introducción

El objetivo de esta práctica esta práctica es aprender sobre los distintos modelos de aprendizaje no supervisado en Python y sus herramientas para interpretar los resultados. Para ello tenemos unos datos sobre accidentes de tráfico con los que tendremos que construir 5 casos diferentes para pasárselos a otros 5 modelos diferentes.

Para construir los casos tendremos que fijar una serie de atributos común para todos, en este caso he seleccionado el total de víctimas, de muertos, de heridos graves y de heridos leves. Esta selección no está hecha aleatoriamente, son los 4 parámetros más importantes a la hora de evaluar la gravedad de un accidente de tráfico.

En cuanto a los modelos elegidos he escogido estos 5:

* KMEANS
* DBSCAN
* BIRCH
* AGGLOMERATIVE
* MEANSHIFT

# Casos de estudio

## Caso 1: colisiones de vehículos con más de un muerto

Los accidentes de vehículos con más transcendencia en medios de comunicación son aquellos donde hay al menos una víctima mortal. Los accidentes de circulación son una de las principales causas de mortalidad en nuestro país, estudiar las situaciones en las diferentes circunstancias de este tipo de accidentes nos ayudaría a crear un plan de choque contra estos.

|  |  |  |  |
| --- | --- | --- | --- |
|  | TIEMPO | Calinski Harabasz | Silhoutte score |
| KMEANS | 0.0167696475982666 | 56.5934806694178 | 0.49178324765558956 |
| DBSCAN | 0.0019948482513427734 | 12.795024502205473 | 0.72834226132662 |
| BIRCH | 0.014960765838623047 | 28.835536567922613 | 0.4636304793763855 |
| AGGLOMERATIVE | 0.001995086669921875 | 44.41665261125605 | 0.4247884098155005 |
| MEANSHIFT | 0.09081029891967773 | 20.510717348034827 | 0.396876202473151 |

## Caso de estudio 2: accidentes sin víctimas mortales

Los accidentes sin muertos son los accidentes más comunes, aunque sean los que menos relevancia mediática tienen. Puede ser interesante realizar un estudio sobre este tema, cualquier estudio serio sobre este tema sería de gran importancia para una compañía de seguros, ya que serán los sucesos que más les reportarán.

|  |  |  |  |
| --- | --- | --- | --- |
|  | TIEMPO | Calinski Harabasz | Silhoutte score |
| KMEANS | 0.03986382484436035 | 44234.6951281487 | 0.8745535617454485 |
| DBSCAN | 3.6393117904663086 | 179421.58672259317 | 0.9976625594820463 |
| BIRCH | 1.3798198699951172 | 3442.1048578227933 | 0.7315154356444701 |
| AGGLOMERATIVE | 7.191510200500488 | 31673.859486562265 | 0.7721860171743123 |
| MEANSHIFT | 230.02169585227966 | 1.0 | 0.9978632478632479 |

## Caso de estudio 3: accidentes con 10 o más víctimas

Los accidentes múltiples son los más peligrosos y donde más perdidas se producen. Estudiar estos siniestros puede ser útil para tener un protocolo sanitario para actuar con velocidad y eficacia.

|  |  |  |  |
| --- | --- | --- | --- |
|  | TIEMPO | Calinski Harabasz | Silhoutte score |
| KMEANS | 0.020753145217895508 | 69.83226450064954 | 0.5205831629916634 |
| DBSCAN | 0.004987955093383789 | 1.0143139478373564 | 0.20771753490098166 |
| BIRCH | 0.011969804763793945 | 1.0143139478373564 | 0.20771753490098166 |
| AGGLOMERATIVE | 0.004985809326171875 | 55.80650959104075 | 0.8237647762501438 |
| MEANSHIFT | 0.20483684539794922 | 64.46033274296263 | 0.5016698833288613 |

## Caso de estudio 4: accidentes con mal tiempo

La peligrosidad al volante crece exponencialmente cuando las condiciones climáticas son adversas.

|  |  |  |  |
| --- | --- | --- | --- |
|  | TIEMPO | Calinski Harabasz | Silhoutte score |
| KMEANS | 0.04387998580932617 | 16096.363656143792 | 0.8235676356058509 |
| DBSCAN | 1.2140283584594727 | 5205.550407452377 | 0.9919336081066589 |
| BIRCH | 0.8695728778839111 | 1737.5349665723434 | 0.8199819184066919 |
| AGGLOMERATIVE | 3.937746286392212 | 16011.244920760882 | 0.7693751561500967 |
| MEANSHIFT | 63.09492516517639 | 1.0 | 0.9934337582078022 |

## Caso de estudio 5: colisión de vehículos

Los siniestros con colisiones de vehículos implican más de un coche y pueden ser bastante graves. Estos accidentes pueden ser los que más víctimas mortales implican.

|  |  |  |  |
| --- | --- | --- | --- |
|  | TIEMPO | Calinski Harabasz | Silhoutte score |
| KMEANS | 0.04819464683532715 | 15142.74615972808 | 0.8907902859972796 |
| DBSCAN | 3.753314733505249 | 5205.550407452377 | 0.9919336081066589 |
| BIRCH | 0.8908498287200928 | 1737.5349665723434 | 0.8199819184066919 |
| AGGLOMERATIVE | 8.904922485351562 | 16011.244920760882 | 0.7693751561500967 |
| MEANSHIFT | 168.80518889427185 | 1.0 | 0.9934337582078022 |

## Análisis de los resultados

La mayoría de los resultados están dentro de lo esperado, pero hay un problema recurrente. Muchos algoritmos se atoran y tardan mucho tiempo devolviendo resultados que son altamente sospechosos de ser falsos o al menos de no ser creíbles. Esto se puede deber a que utiliza bloques de datos demasiado grandes para los ajustes a lo que los he impuesto.

## KNN con nuevos ajustes

|  |  |  |  |
| --- | --- | --- | --- |
|  | TIEMPO | Calinski Harabasz | Silhoutte score |
| Caso 1 | 0.03889632225036621 | 66.70711418943985 | 0.6587894222370997 |
| Caso 2 | 0.2613081932067871 | 124527.52612745189 | 0.9758221269467139 |
| Caso 3 | 0.05634355545043945 | 322.865200729248 | 0.4434469835111024 |
| Caso 4 | 0.22150039672851562 | 26062.859996408795 | 0.9449019781700694 |
| Caso 5 | 0.2005615234375 | 27708.23519113556 | 0.953113030630687 |

NOTA: las gráficas están en el notebook