

TFG

Aarón Casado Monge¹, Juan José Cuadrado Gallego¹

University of Alcalá, Polytechnic School, Computer Science Department, Scientific and Technological Campus, Polytechnic Building. Office: 0243, 28805, Alcalá de Henares, Madrid, Spain

Abstract

Clusterización (qué es y para qué sirve) - Técnicas (para qué sirven) - Paquetes

Keywords: BigData, Statistics, Biostatistics, Data Science, Clustering

1. Introducción

Desde finales del siglo XX se ha considerado que vivimos en la “era de la información”, una etapa caracterizada por el incremento, desarrollo y propa-
gación de emergentes tecnologías de la información
y comunicación que han permitido al ser humano
romper las barreras de la distancia, el tiempo y el
lugar a la hora de comunicarse y compartir infor-
mación; actividades que han sido decisivas en nues-
tra historia [1]. Sin embargo, la era en la que real-
mente vivimos es la “era de los datos”; cada día
se generan más de dos mil quinientos millones de
petabytes¹ de datos provenientes de negocios, cien-
cias, Internet y casi cualquier actividad del día a día
[2] que acaban volcados en redes de ordenadores,
sitios web, bases de datos y otros medios de alma-
cenaje.

Esta gran cantidad de datos proviene de lo com-
puterizada que está la sociedad.

Email addresses: aaron.casado@uah.es (Aarón Casado Monge), jcg@uah.es (Juan José Cuadrado Gallego)

¹Un Petabyte es una unidad de información o almacenamiento de datos equivalente a un cuadrillón de bytes, mil terabytes o un millón de gigabytes. En este caso, es el equivalente a 2.5 quintillones de bytes.

2. Clustering

Clustering o Cluster Analysis, adaptado al español como **Clusterización**, Agrupamiento o Análisis de Grupos es un método de clasificación no supervisada perteneciente a Data Mining, que busca definir, para una característica determinada o Suceso Elemental (SE) ², un conjunto de grupos de observaciones (suceso) ³ con valores cercanos. Estos grupos son los denominados clusters o grupos y permiten a partir de los diferentes sucesos elementales que configuran dicho suceso, asignar dicho SE al mismo. Clustering nos permite definir los valores de cada cluster durante el proceso de clasificación [5].

3. Referencias

- [1] Alberts, D. S., & Papp, D. S. (1997). [The information age: An anthology on its impact and consequences](#). Office of the Assistant Secretary of Defense Washington DC Command and Control Research Program (CCRP).
 - [2] Becoming A Data-Driven CEO — Domo. (2018). Data never sleeps 6.0 <https://www.domo.com/solution/data-never-sleeps-6>
 - [3] Everyman's Science
 - [4] Apuntes JJ Data Science
 - [5] Apuntes JJ clustering
 - Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques (3rd ed., p. 740). 225 Wyman Street, Waltham, MA 02451, USA: Morgan Kaufmann Publishers, Elsevier.
- <https://normas-apa.org/referencias/citar-diccionario/>
<https://www.scribbr.es/detector-de-plagio/>
<https://tablesgenerator.com/>
elsevier dos páginas latex

²Definir suceso elemental

³Definir Suceso