

Actividad Evaluable: Patrones con K-means

Jorge Guerrero Díaz A01411752

José Sebastián Pedrero Jiménez A01703331

Armando Gutiérrez Rojo A01702748

Hipótesis: Pokemon de la misma generación tienen un poder de ataque similar.

¿Crees que estos centros puedan ser representativos de los datos? ¿Por qué?

Los centros nos muestran las tendencias de cada clase por así decirlo, pero al no haber una mezcla o empalme de dichos datos vemos como hay una clara segregación o separación que evita la interrelacionalidad.

¿Cómo obtuviste el valor de k a usar?

Se utilizó el algoritmo de "punto de codo" y se seleccionó el punto en el que se nota el inicio de la curva, siendo este el número 4

¿Los centros serían más representativos si usaras un valor más alto? ¿Más bajo?

Al utilizar el algoritmo de "punto de codo", seleccionamos la distancia media entre los puntos y el centroide cuando la tasa de incremento o descenso se afila. Por lo tanto, si se utilizara un valor más alto de k la distancia media entre los puntos disminuirá y de igual forma si se disminuye el valor de k la distancia media entre los puntos aumentará. Es por esto que si se usa un valor más alto o más bajo, no sería más representativo.

¿Qué distancia tienen los centros entre sí? ¿Hay alguno que esté muy cercano a otros?

Todos los centros se encuentran a una distancia muy parecida. Cada uno se encuentra a una distancia de aproximadamente 25 unidades en el eje de las X y 25 unidades en el eje de las Y del centro anterior.

¿Qué pasaría con los centros si tuviéramos muchos outliers en el análisis de cajas y bigotes?

Implicaría muchas más diferencias o casos que salen fuera de las tendencias generales, haciendo muchas más diferencias entre los valores, e indicando que no hay tendencias o relaciones de algún modo,

¿Qué puedes decir de los datos basándose en los centros?

Los centros nos indican, el promedio o el punto de cada clase donde se centra el valor más relevante por así decirlo, que a raíz de dicho centro los demás datos emanan y se reparten, por lo visto en los centros de nuestra base de datos vemos que hay datos únicos no compartidos que indican que no hay alguna relación entre los demás.

Link del Github:

https://github.com/MandukoGR/TC1002S_Equipo2