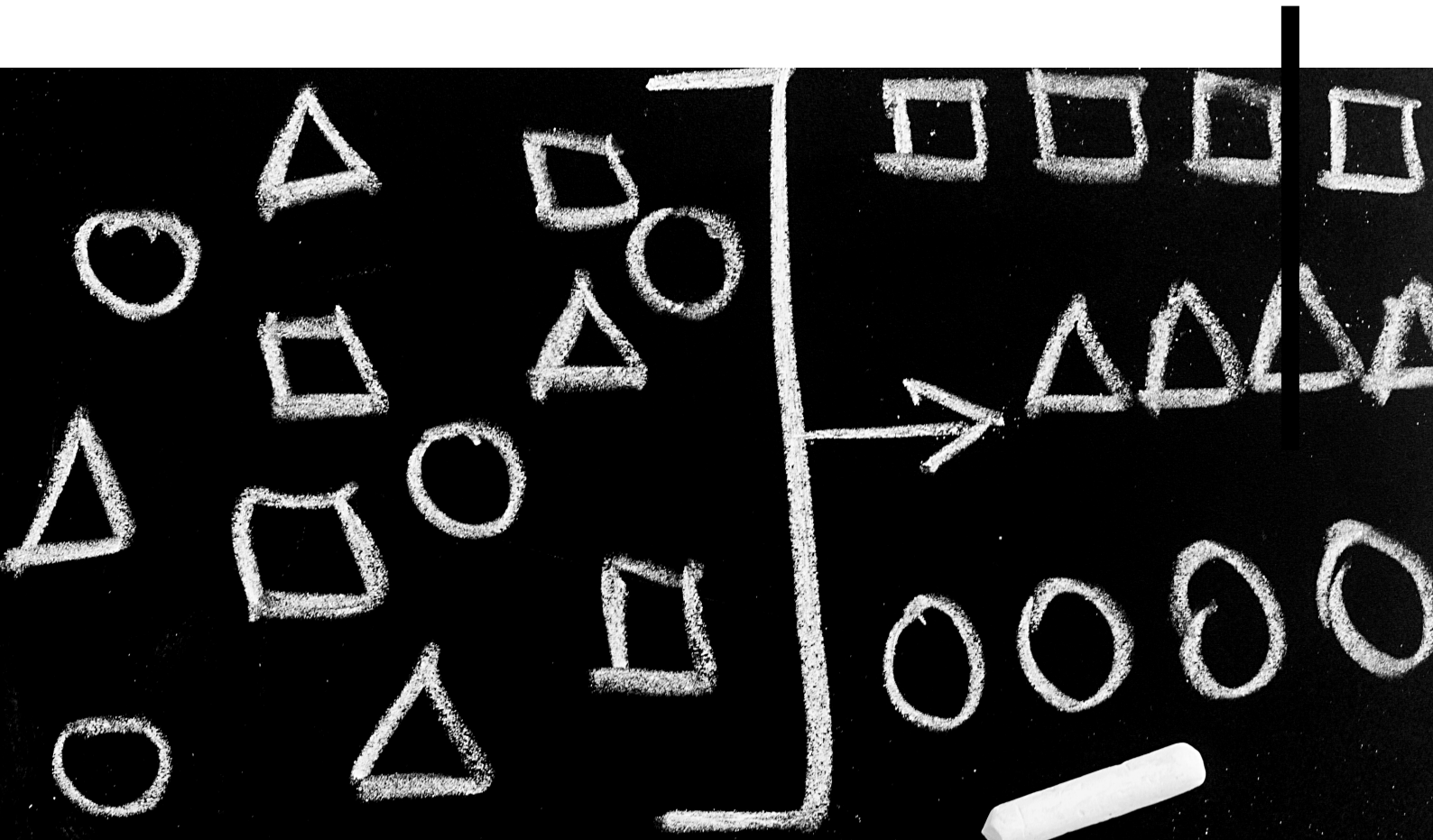


Inteligencia de negocios

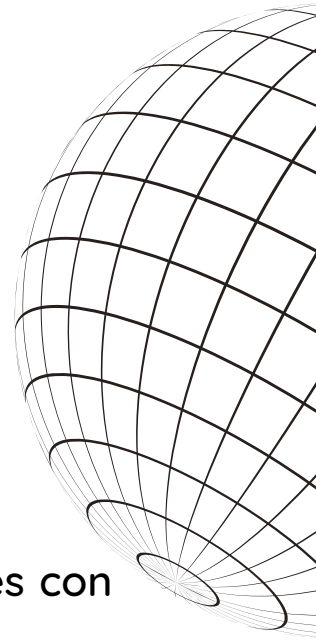
CATEGORIZACIÓN CON K-MEANS

Universidad Modelo

2022



INTRODUCCIÓN



La categorización es el intento progresivo de agrupar la información recogida en base a ciertos criterios, lo que ayuda a la mejor comprensión y construcción de estructuras.

Una manera sencilla de categorizar en programación es con el algoritmo k-means,

K-means es un algoritmo de clasificación no supervisada (clusterización) que agrupa objetos en k grupos basándose en sus características. El agrupamiento se realiza minimizando la suma de distancias entre cada objeto y el centroide de su grupo o cluster.

En este proyecto estaremos tomando una base de datos de un zoológico y estaremos categorizando mediante el algoritmos k-means a los animales según sus características en 7 grupos distintos.

zoo classification

DATA



Para este proyecto utilizaremos la base de datos llamada "Zoo Animal Classification". Este conjunto de datos consiste en una lista de 101 animales de un zoológico en Brasil, en donde tenemos distintas clases de animales con distintas características.

Cada columna de la tabla representa una característica de estos animales como por ejemplo si tienen patas, ponen huevos, si son domésticos o no, etc. Entonces se utilizara el "1" si el animal cumple con la característica y "0" en caso de no cumplirla

Cada columna representa una característica del animal y se utiliza el valor de "1" en caso de cumplir con la característica o "0" en caso de no cumplirla. Se agrego una columna llamada grupo para la clasificación de los animales por tipo: mamíferos, aves, reptiles, peces, anfibios, insectos e invertebrados.

Las características de la colección de datos son las siguientes:

- nombre_animal
- cabello
- plumas
- huevos
- Leche
- aerotransportado
- acuático
- depredador
- dentado
- columna vertebral
- respira
- venenoso
- aletas
- piernas
- cola
- Doméstico
- tamaño de gato

DESARROLLO

Usaremos Python como nuestro lenguaje junto con Pandas y Google Colab como nuestras herramientas para ejecutar el algoritmo en nuestros datos y obtener como resultado nuestros grupos en un formato facilite su interpretación.

Seguiremos los siguientes pasos:

- Importar las librerías que necesitaremos
- Seleccionar nuestro documento de entrada (Base de datos)
- Seleccionar el nombre de la columna que no usaremos.
- Definir el número de grupos con el parámetro "n_clusters", en este caso definimos 7.
- Incrustar nuestra predicción en un nuevo set de datos usando una columna con el nombre "grupo".
- Incrustar la columna del nombre de los animales.
- Exportar el archivo zoo.csv a la carpeta deseada en este caso Output.

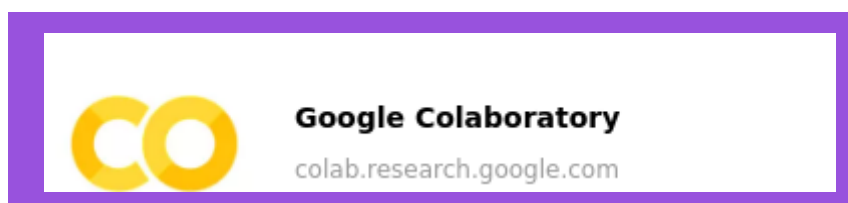
RECOMENDACIONES Y CONSIDERACIONES

Considera que se trata de 7 grupos distintos iniciando desde el número 0. De esta forma podemos estar seguros que los animales comparten características que los hacen pertenecer a un mismo grupo.

Si utilizas google colab recuerda subir la base de datos cada vez se recicle el entorno de ejecución y no olvides agregar la carpeta llamada Output en tus archivos.

Esta información podría ser de utilidad para la toma de decisiones al momento de planificar la distribución de los animales de un zoológico, para que de esta manera sea mas armonioso para los visitantes y los cuidadores.

Proyecto:



RESULTADOS

El primer grupo obtenido son reptiles ovíparos es decir que ponen huevos, aunque la mayoría de los reptiles son ovíparos, hay algunos que suelen ser vivíparos

animal_name	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	backbone	breathes	venomous	fins	legs	tail	domestic	catsize	grupo
turtle	0	0	1	0	0	1	0	0	1	1	0	0	4	1	1	1	0
chameleon	0	0	1	0	0	0	0	1	1	1	0	0	4	1	1	0	0
iguana	0	0	1	0	0	0	1	1	1	1	0	0	4	1	1	1	0
lizard	0	0	1	0	0	0	1	1	1	1	0	0	4	1	0	0	0
gecko	0	0	1	0	0	0	0	1	1	1	0	0	4	1	1	0	0
crocodile	0	0	1	0	0	1	1	1	1	1	0	0	4	1	0	1	0
alligator	0	0	1	0	0	1	1	1	1	1	0	0	4	1	0	1	0
gharial	0	0	1	0	0	1	1	1	1	1	0	0	4	1	0	1	0
skink	0	0	1	0	0	0	0	1	1	1	0	0	4	1	0	0	0
salamander	0	0	1	0	0	1	0	1	1	1	0	0	4	1	1	0	0

Como segundo grupo obtuvimos animales acuáticos, en específico peces.

animal_name	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	backbone	breathes	venomous	fins	legs	tail	domestic	catsize	grupo
anchovy	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	1
flounder	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	1
halibut	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	1
mackerel	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	1
barracuda	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0	1	1
marlin	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	1
trout	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	1

En el grupo 3 obtuvimos insectos que si bien tienen alas todos, no todos vuelan, aunque si la mayoría.

animal_name	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	backbone	breathes	venomous	fins	legs	tail	domestic	catsize	grupo
mosquito	0	0	1	0	1	0	0	0	0	1	0	0	6	0	0	0	2
hornet	0	0	1	0	1	0	1	0	0	1	1	0	6	0	0	0	2
cricket	0	0	1	0	0	0	0	0	0	1	0	0	6	0	0	0	2
beetle	0	0	1	0	0	0	0	0	0	1	0	0	6	0	0	0	2
butterfly	1	0	1	0	1	0	0	0	0	1	0	0	6	0	0	0	2
palmetto	0	0	1	0	1	0	1	0	0	1	0	0	6	0	0	0	2
cockroach	0	0	1	0	0	0	0	0	0	1	0	0	6	0	0	0	2
mantis	0	0	1	0	0	0	1	0	0	1	0	0	6	0	0	0	2
dragonfly	0	0	1	0	1	0	0	0	0	1	0	0	6	0	0	0	2
aphid	0	0	1	0	0	0	0	0	0	1	0	0	6	0	0	0	2
cicada	0	0	1	0	1	0	0	0	0	1	0	0	6	0	0	0	2
antlion	0	0	1	0	1	0	0	0	0	1	0	0	6	0	0	0	2

RESULTADOS

En el grupo cuatro solo se obtuvieron los arácnidos , en este caso solo hay uno.

animal_name	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	backbone	breathes	venomous	fins	legs	tail	domestic	catsize	grupo
spider	0	0	1	0	0	0	1	1	0	1	1	0	8	0	0	0	3

En el grupo 5 también obtuvimos reptiles pero en especifico serpientes

animal_name	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	backbone	breathes	venomous	fins	legs	tail	domestic	catsize	grupo
python	0	0	1	0	0	0	1	1	1	1	1	0	0	1	0	1	4
boa	0	0	1	0	0	0	1	1	1	1	0	0	0	1	0	1	4
adder	0	0	1	0	0	0	1	1	1	1	1	0	0	1	0	1	4

En el grupo 6 obtuvimos animales ovíparos pero esta vez solo los que son del tipo moluscos.

animal_name	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	backbone	breathes	venomous	fins	legs	tail	domestic	catsize	grupo
scallop	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	5
snail	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	5
silkworm	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	5
jellyfish	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	1	5
squid	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	5

Como ultimo grupo obtuvimos de igual manera ovíparos pero ahora en especifico los anfibios.

animal_name	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	backbone	breathes	venomous	fins	legs	tail	domestic	catsize	grupo
siren	0	0	1	0	0	1	0	1	1	1	0	0	2	1	0	0	6
tree frog	0	0	1	0	0	1	0	1	1	1	0	0	4	0	0	0	6
dart frog	0	0	1	0	0	1	0	1	1	1	0	0	4	0	0	0	6
firebelly	0	0	1	0	0	1	0	1	1	1	0	0	4	0	0	0	6
wart toad	0	0	1	0	0	1	0	1	1	1	0	0	4	0	0	0	6

CONCLUSIÓN

En conclusión, podemos utilizar las tecnologías como en este caso machine learning para facilitar tareas diarias pero siempre tomando en cuenta que debe intervenir una persona experta, tanto en tecnología como en el área en la que se trabaja.



INSPIRACIÓN

Este proyecto fue parte de la materia de inteligencia de negocios impartida por el profesor Luis Novelo en la Universidad Modelo de Mérida.

LIBRERIAS

- joblib: 1.0.1
- numpy 1.21.0
- pandas: 1.3.0
- scikit-learn: 0.24.2
- scipy: 1.7.0

REFERENCIAS

<https://github.com/PhinanceScientist/Candy-Clustering>

<https://www.sciencedirect.com/science/article/pii/S1877050915035929>

<https://courses.lumenlearning.com/wm-retailmanagement/chapter/layout-of-products/>

<https://www.zoho.com/inventory/guides/what-is-item-grouping.html>

<https://erply.com/why-the-right-product-selection-is-crucially-important-for-a-retailer/>