

ANÁLISIS DE ALGORITMOS DE APRENDIZAJE  
AUTOMÁTICO APLICADOS A LA  
CLASIFICACIÓN DE CUERPOS CELESTES:  
ESTRELLAS, QUÁSARES Y GALAXIAS.

Instituto Tecnológico y de Estudios Superiores de Monterrey.

Participantes:

Dulce Nahomi Bucio Rivas.

Emiliano Vivas Rodríguez.

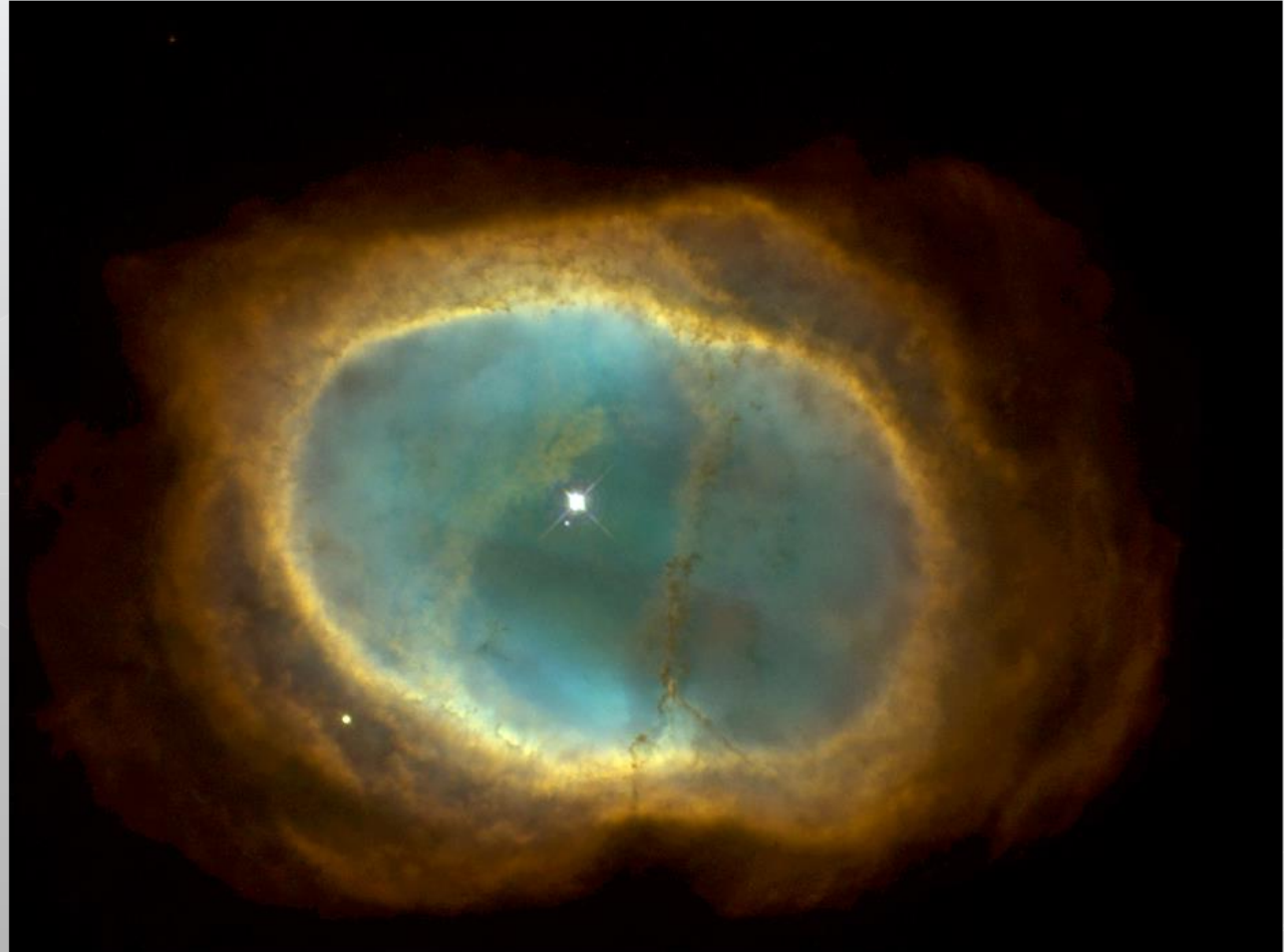
Regina Alexia Blas Flores.

Atoany Nazareth Fierro Radilla.

# INTRODUCCIÓN

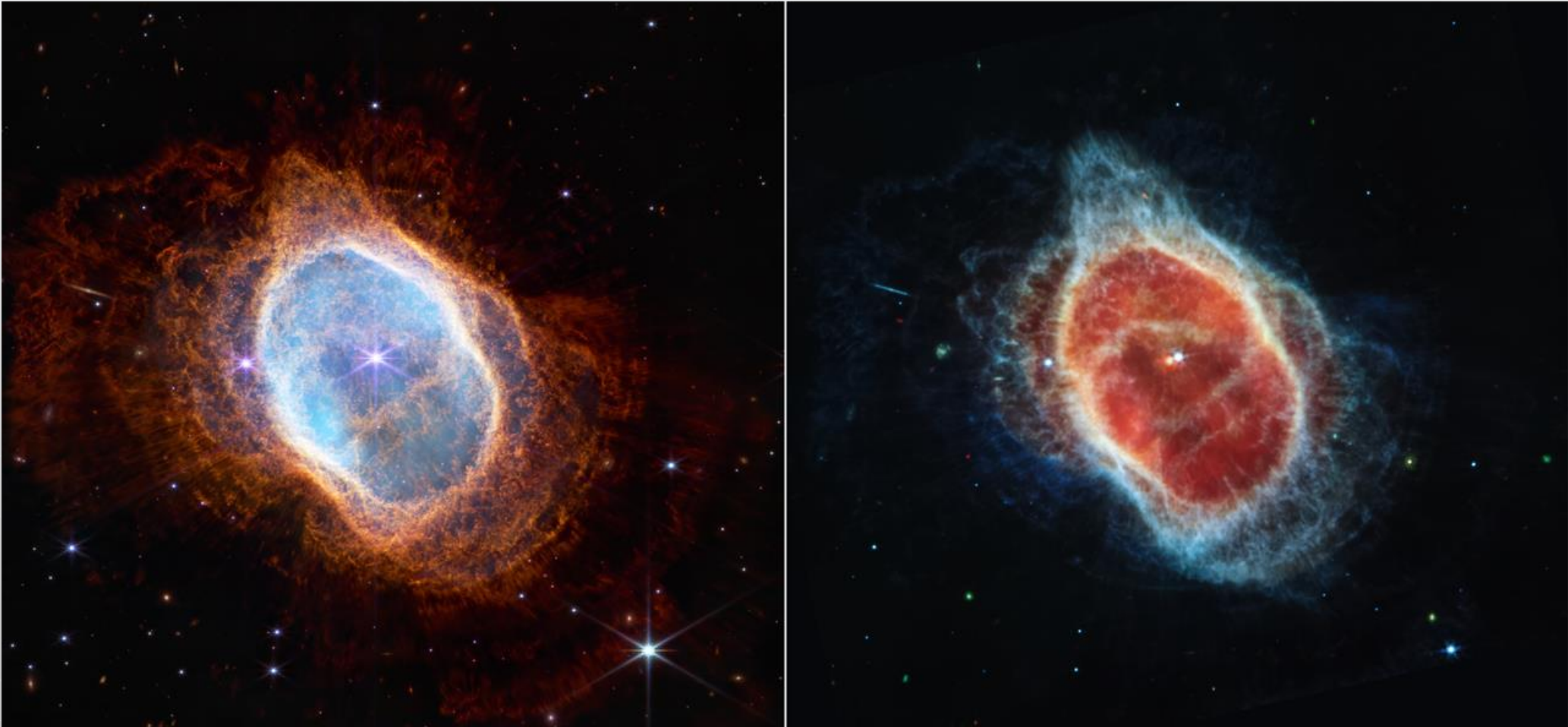
Los avances científicos y tecnológicos resultan imprescindibles para la obtención de los avances en el área de la exploración espacial y el conocimiento documentado sobre el Universo.

La era de la información, la era del BigData, ha permitido realizar incontables descubrimientos a través de la ciencia de datos, el *Machine Learning* y la inteligencia artificial.



NGC 3132, Nebulosa del Anillo del Sur, capturada por el telescopio espacial Hubble.

# INTRODUCCIÓN



NGC 3132, Nebulosa del Anillo del Sur, capturada por el telescopio espacial James Webb.



# INTRODUCCIÓN

The Sloan Digital Sky Survey se trata de un proyecto de investigación del espacio mediante imágenes en el espectro visible y de corrimiento al rojo.

Ha creado los mapas tridimensionales con imágenes multicolores profundas de un tercio del cielo y espectros de más de tres millones de objetos astronómicos.

Las observaciones han sido capturadas por el telescopio *Sloan Digital Sky Survey V* de ángulo amplio y de 2.5 metros situado en el observatorio Apache Point de Nuevo México, Estados Unidos de América.

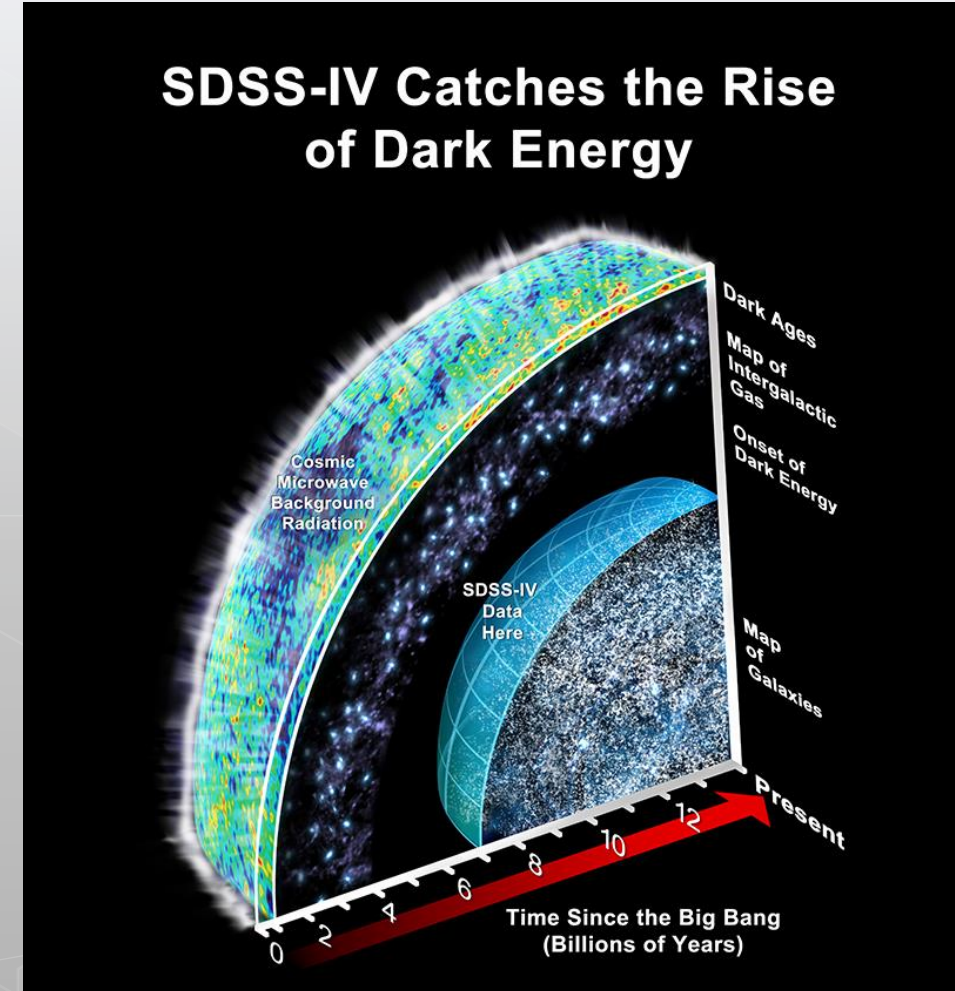


Vista del telescopio *Sloan Digital Sky Survey V*.

# METODOLOGÍA

Sloan Digital Sky Survey ha permitido la difusión pública de grandes volúmenes de datos a través de sus bases de datos, los SDSS Data Release 16 y 17. Las publicaciones de datos SDSS son acumulativas, por lo que DR17 incluye toda la cobertura del cielo de publicaciones anteriores, esto desde julio de 2014. Estas colecciones incluyen 6 tipos de observaciones:

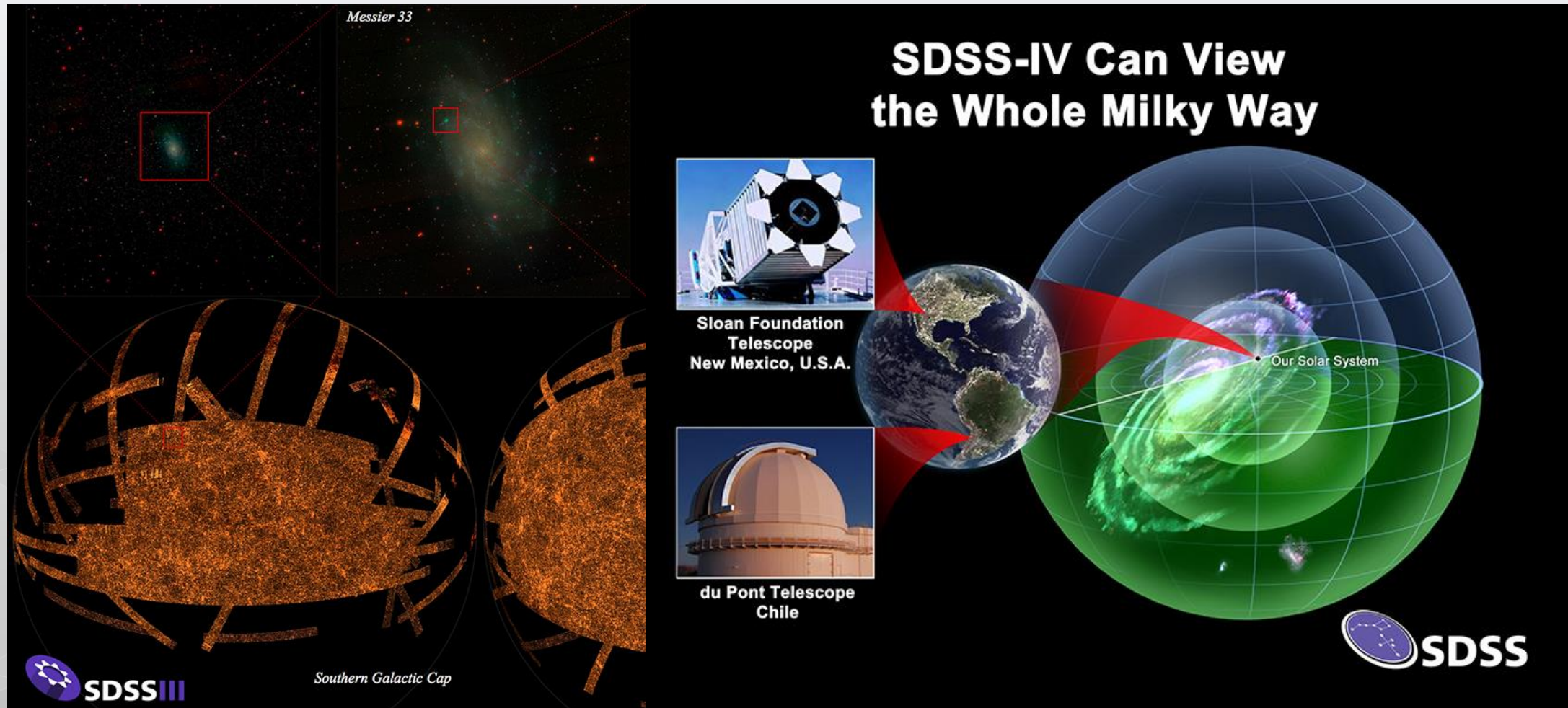
- Imágenes.
- Espectros ópticos (SDSS, SEGUE, BOSS, SEQUELS, eBOSS).
- Espectros infrarrojos (APOGEE, APOGEE-2).
- Espectros IFU (MaNGA).
- Espectros de biblioteca estelar (MaStar).
- Catálogo de datos (parámetros medidos a partir de imágenes y espectros, como magnitudes y corrimientos al rojo).



*The Extended Baryon Oscillation Spectroscopic Survey (eBOSS).*



# METODOLOGÍA

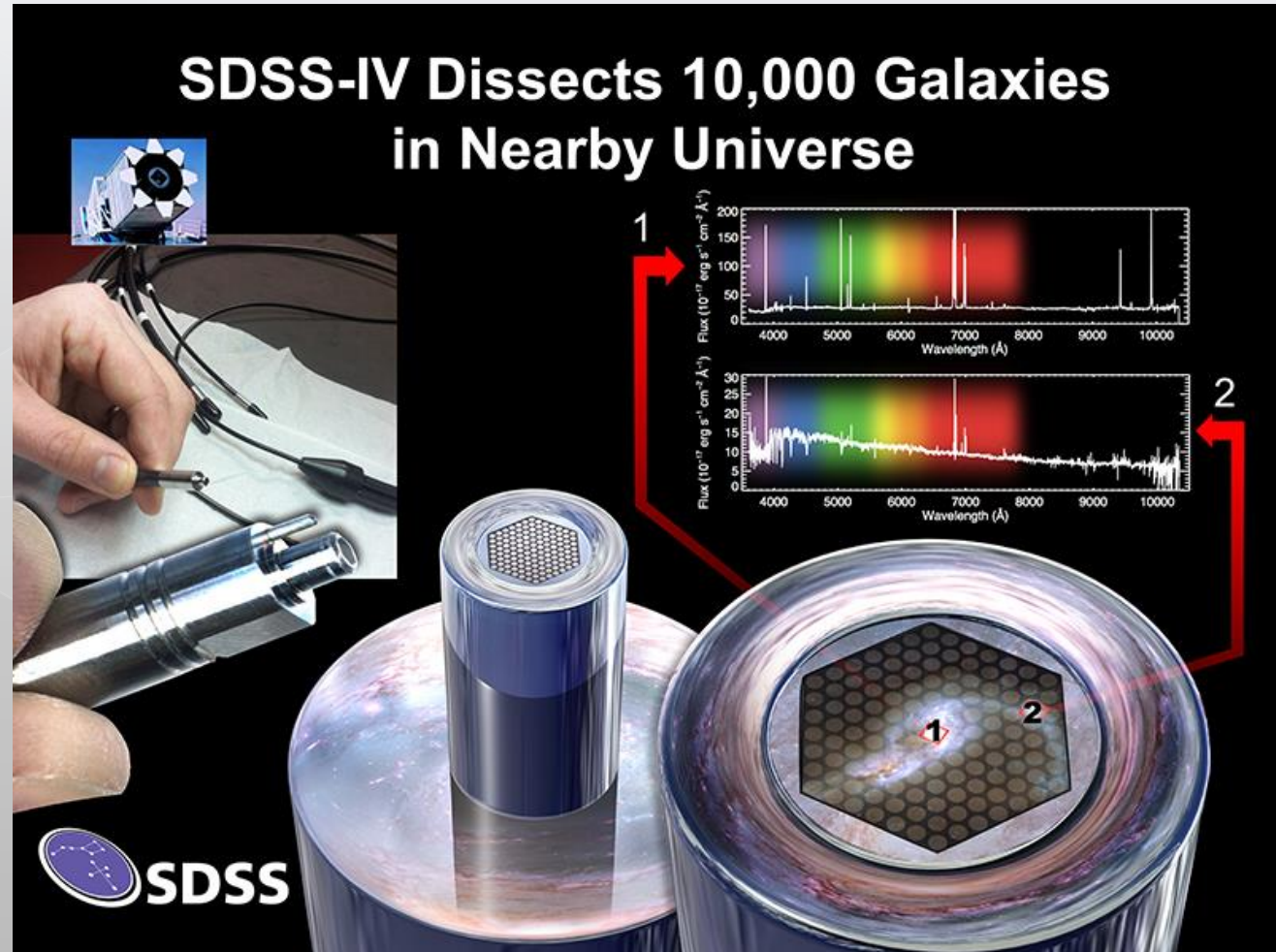


*Sloan Digital Sky Surveys, technology for mapping the Universe.*

# METODOLOGÍA

La base de datos del SDSS Data Release 16 y 17 incluye 18 atributos, de los cuales, 11 campos resultaron útiles:

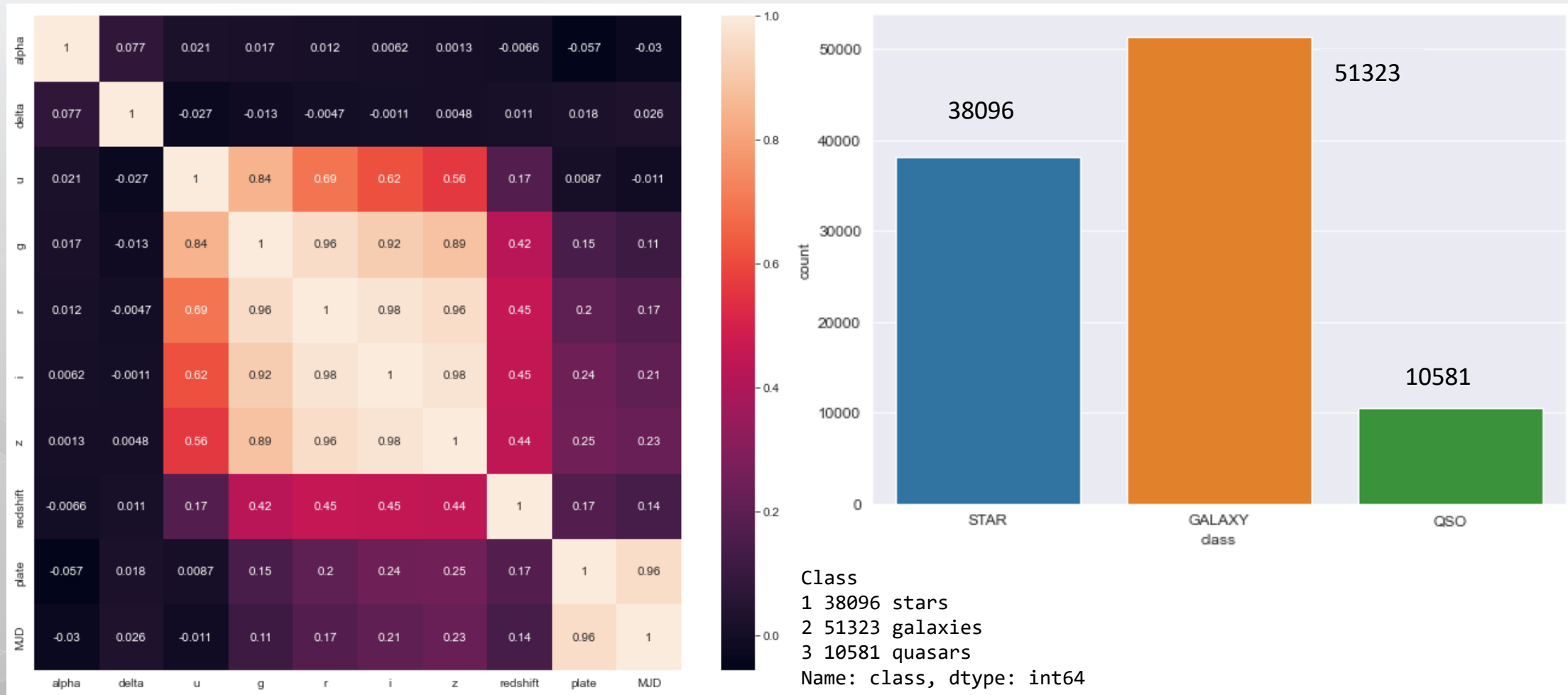
- |                    |                      |
|--------------------|----------------------|
| 1. right_ascension | 7. z_magnitude       |
| 2. Declination     | 8. Class             |
| 3. u_magnitude     | 9. Redshift          |
| 4. g_magnitude     | 10. plate_id         |
| 5. r_magnitude     | 11. observation_date |
| 6. i_magnitude     |                      |



*Mapping Nearby Galaxies at APO (MaNGA).*



# METODOLOGÍA



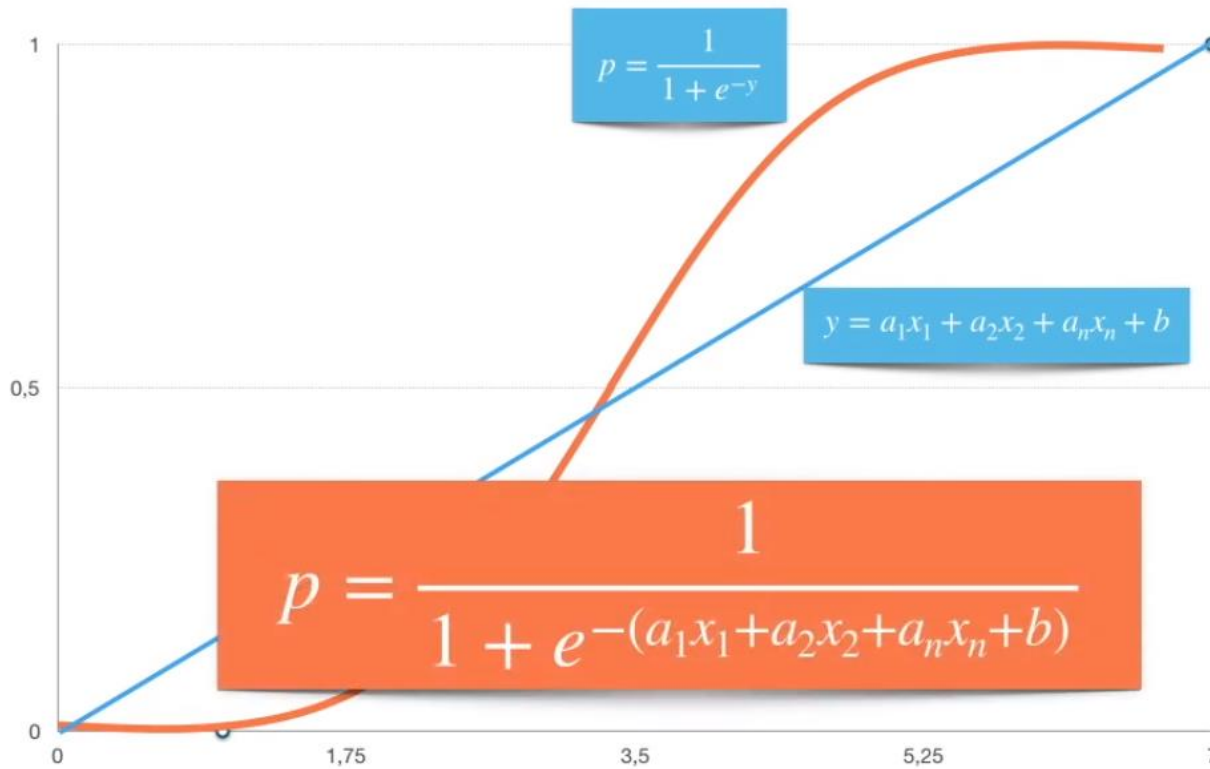
Correlación de variables en la base de datos.

Frecuencia de clases en la base de datos del SDSS 16 y 17.



# RESULTADOS Y ANÁLISIS

## Regresión Logística.



4.557223558425903

	precision	recall	f1-score	support
1	0.97	1.00	0.99	11463
2	0.99	0.97	0.98	15366
3	0.96	0.96	0.96	3171
accuracy			0.98	30000
macro avg	0.97	0.98	0.98	30000
weighted avg	0.98	0.98	0.98	30000

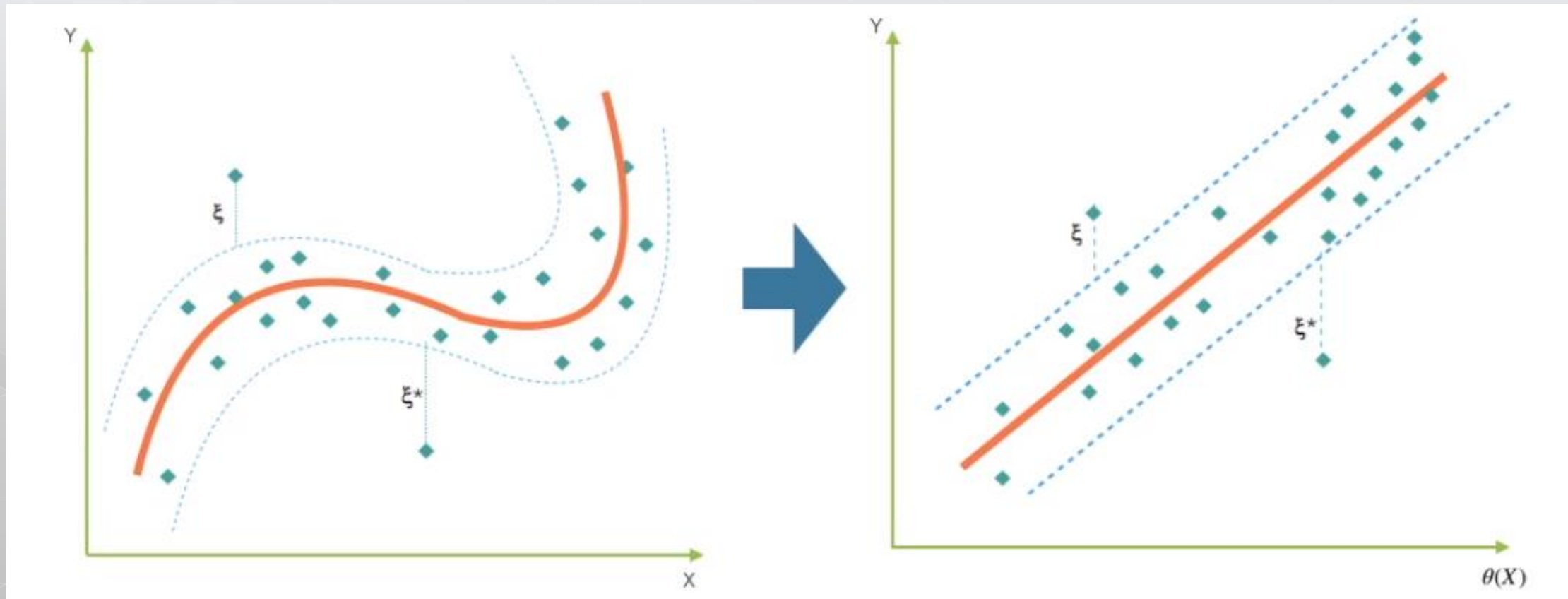
	0	1	2
0	11453	0	10
1	295	14949	122
2	5	124	3042

Accuracy: 0.98

Datos obtenidos tras la Regresión Logística.

# RESULTADOS Y ANÁLISIS

Máquinas Vectoriales de Soporte.



Máquinas Vectoriales de Soporte, algoritmo con aplicación en *Machine Learning*.

# RESULTADOS Y ANÁLISIS

27.514996767044067

	precision	recall	f1-score	support
1	0.98	1.00	0.99	11463
2	0.99	0.98	0.99	15366
3	0.98	0.95	0.96	3171
accuracy			0.99	30000
macro avg	0.98	0.98	0.98	30000
weighted avg	0.99	0.99	0.99	30000

	0	1	2
0	11463	0	0
1	190	15119	57
2	5	165	3001

Kernel: linear.  
Accuracy: **0.99**

167.49528312683105

	precision	recall	f1-score	support
1	0.97	0.80	0.87	11463
2	0.85	0.98	0.91	15366
3	0.99	0.90	0.94	3171
accuracy			0.90	30000
macro avg	0.94	0.89	0.91	30000
weighted avg	0.91	0.90	0.90	30000

	0	1	2
0	9115	2348	0
1	283	15051	32
2	3	310	2858

Kernel: polynomial.  
Accuracy: **0.90**

Datos obtenidos tras la aplicación de Máquinas Vectoriales de Soporte.



# RESULTADOS Y ANÁLISIS

16.303219318389893

	precision	recall	f1-score	support
1	0.98	1.00	0.99	11463
2	0.99	0.98	0.98	15366
3	0.99	0.93	0.96	3171
accuracy			0.98	30000
macro avg	0.98	0.97	0.98	30000
weighted avg	0.98	0.98	0.98	30000

	0	1	2
0	11459	4	0
1	273	15048	45
2	5	208	2958

Kernel: RBF, Radial Basis Function.  
Accuracy: **0.98**

48.17405891418457

	precision	recall	f1-score	support
1	0.84	0.87	0.85	11463
2	0.80	0.78	0.79	15366
3	0.54	0.51	0.52	3171
accuracy			0.79	30000
macro avg	0.73	0.72	0.72	30000
weighted avg	0.79	0.79	0.79	30000

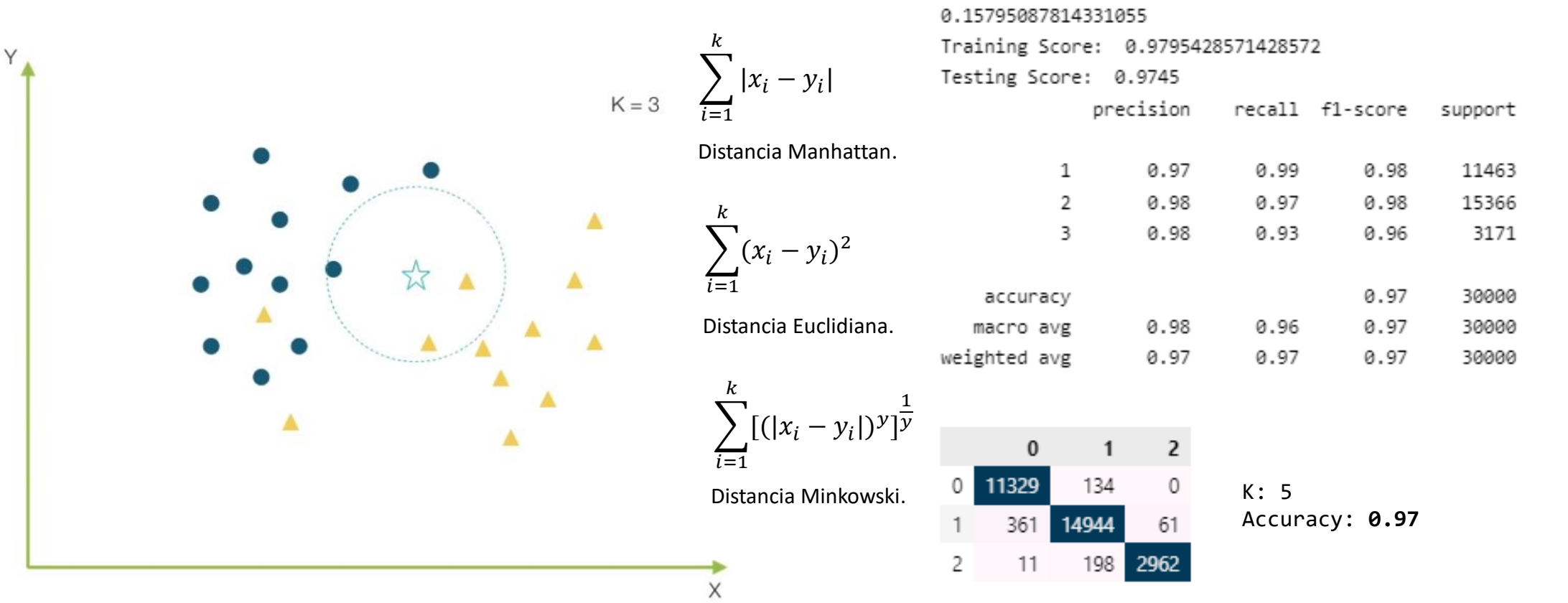
	0	1	2
0	10004	1459	0
1	1920	12055	1391
2	49	1505	1617

Kernel: sigmoid.  
Accuracy: **0.79**

Datos obtenidos tras la aplicación de Máquinas Vectoriales de Soporte.

# RESULTADOS Y ANÁLISIS

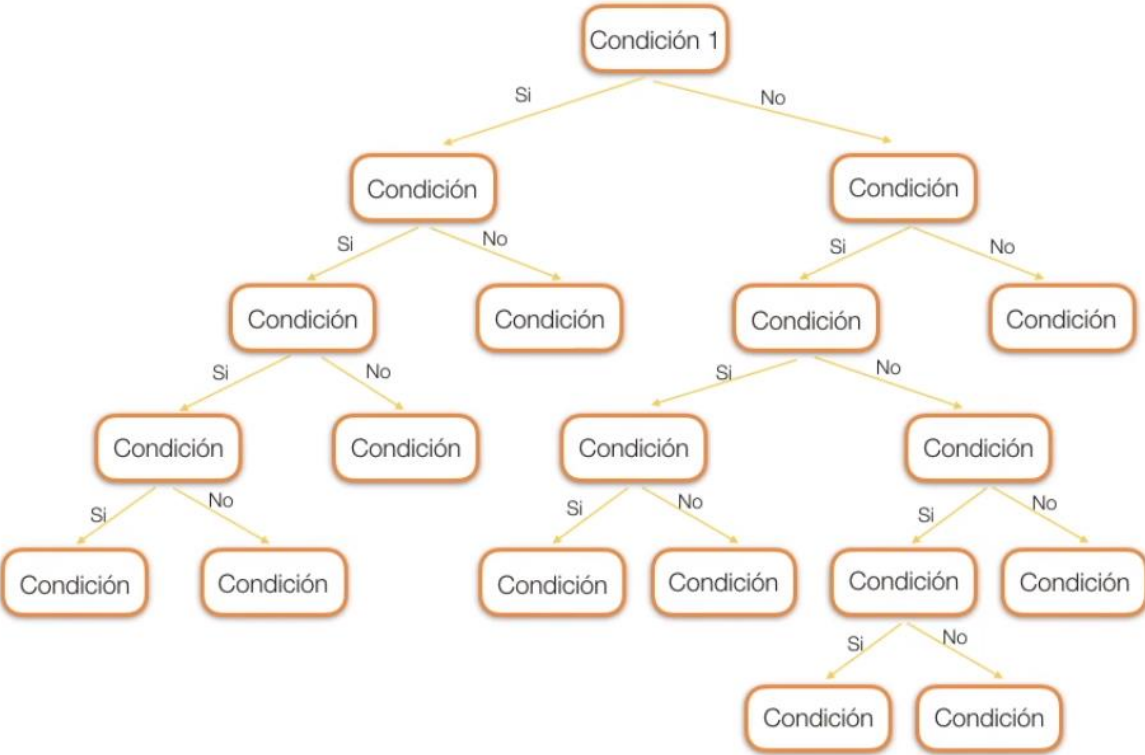
## K Vecinos Más Cercanos.



Datos obtenidos tras la aplicación de K Vecinos Más Cercanos.

# RESULTADOS Y ANÁLISIS

## Árboles de Decisión.



0.7910463809967041  
Training Score: 0.9955  
Testing Score: 0.9897

	precision	recall	f1-score	support
1	1.00	1.00	1.00	11463
2	0.99	0.99	0.99	15366
3	0.98	0.94	0.96	3171
accuracy			0.99	30000
macro avg	0.99	0.98	0.98	30000
weighted avg	0.99	0.99	0.99	30000

	0	1	2
0	11434	29	0
1	14	15278	74
2	4	188	2979

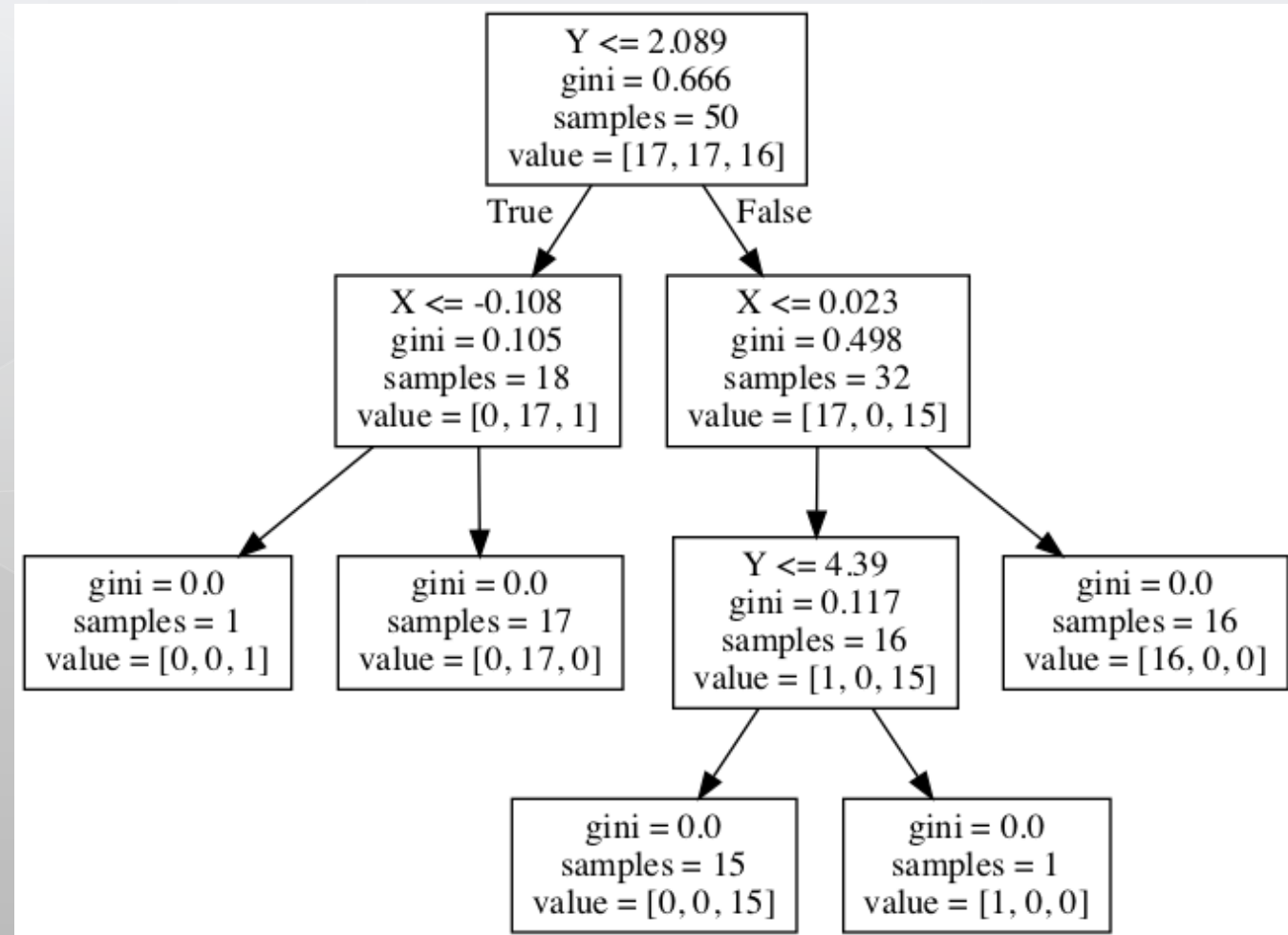
Random state: 0  
Max depth: 10  
Criterion: entropy  
Accuracy: 0.99

Datos obtenidos tras la aplicación de Árboles de Decisión.



# CONCLUSIONES

Pese a contar con un truncamiento en el nivel 10 de bifurcaciones, los árboles de decisión implementados en las colecciones de *SDDS Data Releases* 16 y 17 , a diferencia de otros métodos computacionales, resulta ser el mejor algoritmo de *Machine Learning* que ha permitido clasificar estrellas, quásares y galaxias con un menor margen de error en el proceso, una exactitud del 0.99, un coeficiente de precisión significativo.



Los Árboles de Decision resulta ser el algoritmo de *Machine Learning* supervisado con mayor precisión para clasificar cuerpos celestes.

# Fuentes de información

- "SDSS Data Releases". <https://www.sdss.org/science/data-release-publications> (accedido el 14 de octubre de 2022).
- SDSS. <https://www.sdss.org/> (accedido el 14 de octubre de 2022).
- "Southern Ring Nebula (NIRCam and MIRI Images Side by Side)". WebbTelescope.org  
<https://webbtelescope.org/contents/media/images/2022/033/01G709QXZPFH83NZFAFP66WVCZ?keyword=ngc%203132> (accedido el 14 de octubre de 2022).
- "IMAGES". HubbleSite.org. <https://hubblesite.org/resource-gallery/images> (accedido el 14 de octubre de 2022).
- "scikit-learn: machine learning in Python — scikit-learn 1.1.2 documentation". scikit-learn: machine learning in Python — scikit-learn 0.16.1 documentation.  
<https://scikit-learn.org/stable/> (accedido el 14 de octubre de 2022).