

# CIBERTIC<sup>2025</sup>

Por CUDI

Congreso Internacional de Ciberseguridad,  
Tecnologías, Innovación y Ciencia

**19 - 22 MAYO**

**GUADALAJARA, MÉXICO**

**Hotel Barceló**

Revisión de algunos algoritmos de Machine Learning aplicados a la clasificación de espectros estelares

MC Areli Pérez Aparicio – Dra. Silvana Navarro – Dr. Luis Corral

# ÍNDICE

- Introducción
- Antecedentes
- Gaia Data Release 3 (Gaia DR3)
- Propósito de Gaia DR3
- Algunas investigaciones anteriores
- Algoritmos de clasificación analizados
- Resultados
- Conclusiones

# Introducción

Desde sus inicios, la astronomía ha sido mucho más que la contemplación del cielo; ha servido para medir el tiempo, y ha actuado como un motor para el avance tecnológico, empujando los límites del ingenio humano.

Tales como:

- Dispositivos de carga acoplada
- GPS

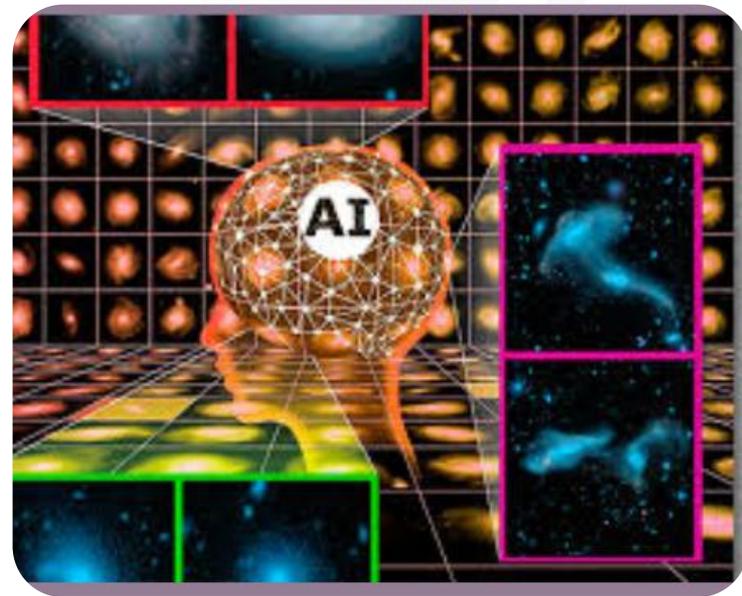
# Introducción

El desafío que representa el manejo de grandes volúmenes de datos previstos para proyectos como el Hubble, Gaia y, más recientemente, el James Webb Space Telescope (JWST), ha dado origen a avances en técnicas computacionales automáticas.



# Introducción

La tercera entrega de datos de Gaia (DR3) ofrece una base de datos sin precedentes de espectros estelares, crucial para entender la composición, evolución y propiedades físicas de las estrellas. Entonces, aplicar algoritmos de Machine Learning en esta base de datos permite unir la Astronomía a las Ciencias Computacionales.



# Antecedentes

El estudio del universo se ha basado históricamente en la recopilación y análisis de datos astronómicos, los cuales han permitido desarrollar modelos sobre la estructura, evolución y composición de los cuerpos celestes. Con el avance de la tecnología, la cantidad de datos disponibles ha crecido exponencialmente, impulsando el uso de nuevas técnicas computacionales para su procesamiento e interpretación.



# Antecedentes

- **Espectro:** proporciona información detallada sobre la composición química, temperatura y dinámica del objeto observado.
- **Datos astronómicos:** son el resultado de observaciones realizadas mediante telescopios ópticos, infrarrojos, de radio, etc.; ya sea en Tierra o en el espacio.
- **Datos fotométricos:** miden la cantidad de luz recibida de un objeto en distintas bandas del espectro electromagnético.
- **Datos espectroscópicos:** contienen información detallada sobre la distribución de la luz en función de la longitud de onda.
- **Datos astrométricos:** incluyen mediciones precisas de la posición, movimiento propio y paralaje de los objetos celestes.

# Gaia Data Release 3 (Gaia DR3)

Es una publicación de datos astronómicos realizada por la misión espacial Gaia de la Agencia Espacial Europea (ESA). Este catálogo, publicado el 13 de junio de 2022, contiene información detallada de aproximadamente 1.800 millones de objetos celestes, incluyendo estrellas, galaxias y cuerpos del sistema solar.



# Propósito de Gaia DR3

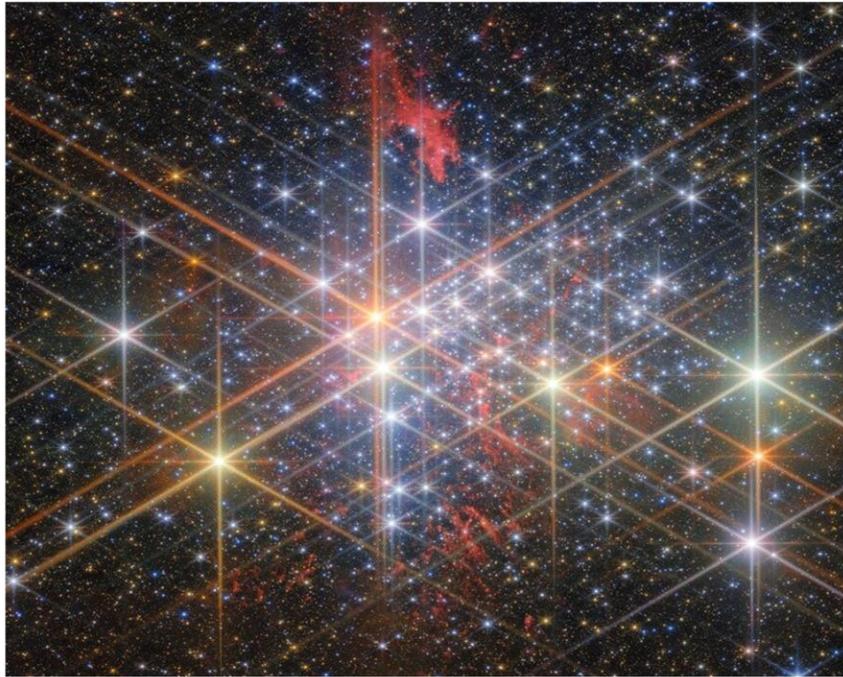
Gaia DR3 amplía y mejora las entregas anteriores al incluir:

- **Astrometría mejorada:** posiciones, paralajes y movimientos propios de los objetos con mayor precisión.
- **Fotometría avanzada:** datos de brillo en múltiples bandas espectrales.
- **Espectroscopía:** espectros de baja resolución y velocidades radiales para una amplia gama de estrellas.
- **Clasificaciones y parámetros estelares:** información sobre las propiedades físicas y químicas de las estrellas, como temperaturas superficiales, masas y edades.

# Algunas investigaciones anteriores

- **Estudio de la estructura de la Vía Láctea:** utilizando datos astrométricos y fotométricos de Gaia DR2, se han mapeado subestructuras y corrientes estelares, proporcionando información sobre la formación y evolución de nuestra galaxia.
- **Detección de exoplanetas:** las mediciones precisas de posiciones y movimientos propios han permitido identificar perturbaciones en estrellas que sugieren la presencia de planetas extrasolares.
- **Caracterización de cúmulos estelares:** con los datos de Gaia, se han redefinido las distancias, edades y composiciones de numerosos cúmulos abiertos y globulares, mejorando la comprensión de la evolución estelar.

# Algoritmos de clasificación analizados

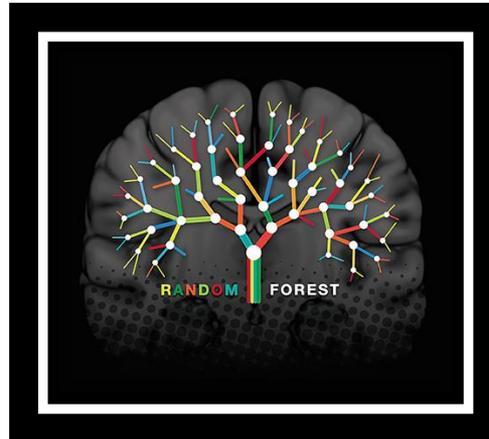


Tras revisar diversos estudios sobre la aplicación de algoritmos de aprendizaje automático en la clasificación de espectros estelares, es posible establecer una comparación objetiva en cuanto a desempeño, ventajas y limitaciones de cada técnica utilizada.

# Algoritmos de clasificación analizados

## *Random Forest (RF):*

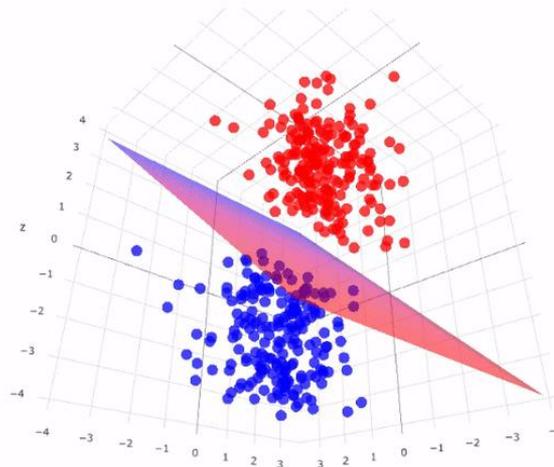
- Este algoritmo ha demostrado ser robusto en tareas de clasificación astronómica, especialmente cuando se dispone de conjuntos de datos balanceados. Su desempeño es destacable en términos de precisión y capacidad para manejar grandes volúmenes de datos. No obstante, en situaciones con datos desbalanceados puede presentar sesgo hacia clases mayoritarias, disminuyendo su efectividad.



# Algoritmos de clasificación analizados

*Support Vector Machine (SVM):*

- SVM muestra buenos resultados cuando los datos están adecuadamente preprocesados y balanceados. Sin embargo, su rendimiento se degrada en presencia de clases desbalanceadas o con ruido, y su complejidad computacional aumenta al trabajar con grandes conjuntos de datos, como los de Gaia DR3.



# Algoritmos de clasificación analizados

## *Artificial Neural Networks (ANN):*

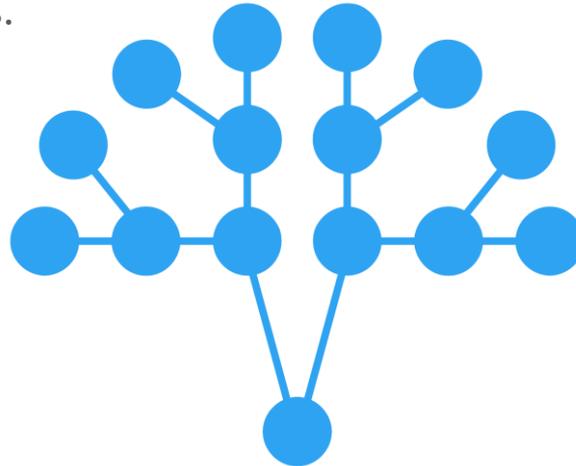
- Las redes neuronales artificiales se posicionan como una de las técnicas más precisas, tanto en conjuntos de datos balanceados como desbalanceados. Su capacidad de aprendizaje y generalización ha sido superior a otros modelos, aunque requieren mayor capacidad de cómputo y tiempo de entrenamiento.



# Algoritmos de clasificación analizados

## *Gradient Boosting (GB):*

- Este algoritmo ha mostrado un rendimiento comparable e incluso superior al de *Random Forest* en varias métricas. Su principal ventaja radica en su capacidad para optimizar errores residuales en forma iterativa. No obstante, también es susceptible al desbalance de clases, por lo que requiere estrategias de preprocesamiento adicionales.



# Algoritmos de clasificación analizados

## *Naive Bayes:*

- Aunque es eficiente en términos de velocidad y simplicidad, su desempeño es limitado debido a la suposición de independencia entre variables, lo cual rara vez se cumple en datos espectrales astronómicos. Como resultado, este algoritmo ha mostrado menor precisión y fiabilidad frente a técnicas más modernas.



## Resultados

- ANN y Gradient Boosting muestran la mayor precisión en clasificación.
- Random Forest y aprendizaje activo optimizan modelos con menos datos etiquetados.
- El Machine Learning supera métodos tradicionales en eficiencia y velocidad.



# Conclusiones

- Machine Learning mejora la clasificación de espectros en Gaia DR3.
- ANN y Gradient Boosting destacan en conjuntos balanceados.
- **Próximos pasos:** implementar aprendizaje profundo para mayor precisión.

# CIBERTIC<sup>20</sup><sub>25</sub>

Por CUDI

Congreso Internacional de Ciberseguridad,  
Tecnologías, Innovación y Ciencia

**19 - 22 MAYO**

**GUADALAJARA, MÉXICO**

**Hotel Barceló**

**[www.cibertic.mx](http://www.cibertic.mx)**

**MC Areli Pérez Aparicio**

**[areli.perez1173@alumnos.udg.mx](mailto:areli.perez1173@alumnos.udg.mx)**