

Comparación de Modelos de IA para la Clasificación de Cuerpos Celestes

Vanessa Carolina Núñez Ramírez
Software Engineer
contact@vanessanunez.pro

Luis Stanley Jovel Portal
Graduate AI Researcher
contact@stanleyjovel.com

Abstract—Este estudio explora la eficacia de dos enfoques de Inteligencia Artificial (IA) para la clasificación precisa de cuerpos celestes. Se utilizaron 100,000 observaciones con 17 características, clasificadas como estrella, galaxia o quasar. Implementamos un modelo de Support Vector Machine (SVM) y una Red Neuronal (NN). Los resultados revelan un alto rendimiento en ambos modelos, con una precisión del 95.36% para la SVM y del 96.70% para la NN. Aunque la NN presenta una ligera ventaja en precisión, la SVM destaca por su eficiencia computacional, siendo más rápida en entrenamiento y prueba, y requiriendo menos memoria. Estos hallazgos sugieren el potencial de la IA para automatizar la identificación de objetos celestes en grandes encuestas astronómicas, con la SVM como una opción particularmente atractiva debido a su eficiencia. El Código usado en este artículo puede ser encontrado en <https://bit.ly/stellar-classification-paper>

I. INTRODUCCIÓN

La clasificación precisa de cuerpos celestes es esencial para avanzar en nuestra comprensión del universo [1]. Tradicionalmente, esta tarea ha sido realizada por astrónomos humanos, pero el volumen y la complejidad de los datos astronómicos actuales presentan un desafío significativo. En este contexto, la Inteligencia Artificial (IA) emerge como una herramienta poderosa para automatizar y optimizar la clasificación de objetos celestes [2]. Este trabajo compara la precisión de dos modelos de IA populares: Support Vector Machine (SVM) y Redes Neuronales (NN), utilizando datos fotométricos y espectroscópicos obtenidos del Sloan Digital Sky Survey (SDSS) [3].

La implementación de técnicas de IA en astronomía no solo promete mejorar la precisión de la clasificación de objetos celestes, sino que también permite el análisis de grandes volúmenes de datos de manera más rápida y eficiente. Las observaciones modernas generan enormes cantidades de datos que son inviables para el análisis manual. Al aplicar modelos de IA, podemos procesar y clasificar estos datos a una velocidad y escala sin precedentes, facilitando descubrimientos astronómicos y contribuyendo al conocimiento científico de manera significativa. En este estudio, se investigan las capacidades de la SVM y las NN en la clasificación de cuerpos celestes, evaluando su rendimiento y eficiencia para determinar su aplicabilidad en el campo de la astronomía.

II. METODOLOGÍA

A. Tratamiento del Dataset

Utilizamos un conjunto de 100,000 observaciones fotométricas que incluye mediciones en múltiples bandas espectrales, así como desplazamientos al rojo, también conocido como corrimiento al rojo, junto con las clasificaciones de los objetos celestes capturados por el SDSS. Cada observación está descrita por 17 características y una columna de clasificación que identifica el objeto como estrella, galaxia o quasar (Tabla I).

TABLE I
TAMAÑO DE LA MUESTRA

Objeto	Num. ejemplos
Galaxia	59,445
Estrella	21,594
Quasar	18,691

Para la clasificación, se seleccionaron solamente las características relevantes para el adecuado entrenamiento y rendimiento de los modelos, descritas en la Tabla II.

TABLE II
CARACTERÍSTICAS UTILIZADAS PARA CLASIFICACIÓN

Feature	Descripción
u	Ultravioleta
g	Azul
r	Rojo
i	Infrarojo
z	Infrarrojo cercano
Redshift	También conocido como corrimiento al rojo

A dichas features se le aplicó un preprocesamiento de normalización eliminando la media y escalando a una varianza unitaria. El puntaje estándar de una muestra x se calcula como:

$$z = \frac{x - \mu}{s}$$

Finalmente, dividimos los datos en conjuntos de entrenamiento y prueba usando la regla del 80/20, reservando el 80% de los datos para entrenamiento y el 20% para pruebas.

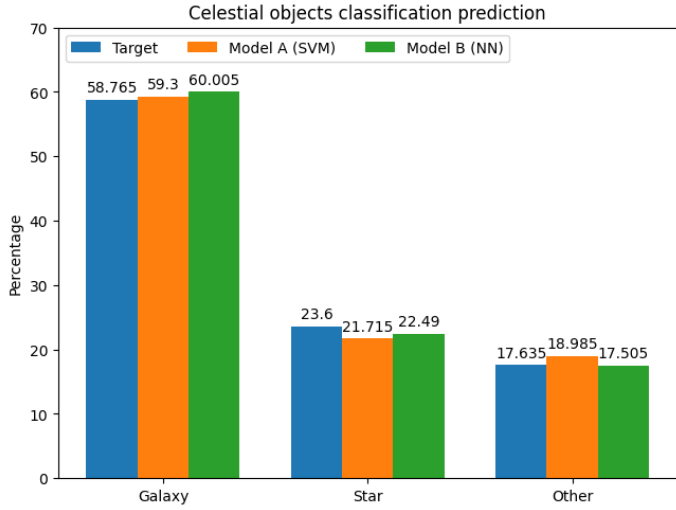


Fig. 1. Tiempo necesario para entrenar el modelo SVM vs Neural Network

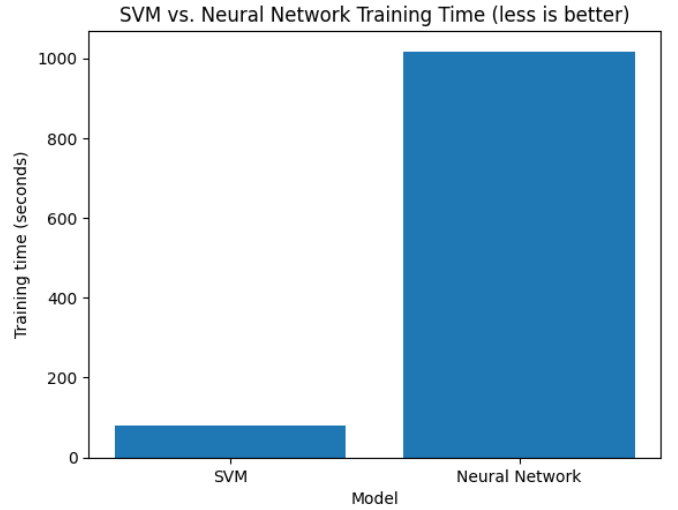


Fig. 2. Tiempo necesario para entrenar el modelo SVM vs Neural Network

B. Modelos

Para el primer modelo, implementamos una SVM con un kernel radial (RBF) y un parámetro de regularización C igual a 100. Luego, entrenamos y evaluamos el modelo utilizando la precisión como métrica de rendimiento.

Para el segundo modelo, utilizamos una NN estándar, configurada con capas densamente conectadas. Entrenamos la red utilizando los datos de entrenamiento y evaluamos su rendimiento en el conjunto de prueba.

III. RESULTADOS

Ambos modelos alcanzan alta precisión (Tabla III), sugiriendo su utilidad potencial en aplicaciones astronómicas prácticas, como la identificación automatizada de objetos en grandes encuestas astronómicas. La ligera ventaja de la NN puede atribuirse a su capacidad para aprender características complejas y jerárquicas de los datos, lo que puede ser beneficioso en la clasificación de objetos astronómicos con estructuras variadas (Figura 1). Sin embargo, el tiempo necesario para entrenar el modelo SVM es considerablemente menor comparado con la NN (92% más rápido) (Figura 2), sugiriendo que el modelo SVM puede ser más adecuado para situaciones en las que el tiempo de entrenamiento es una limitación crítica.

Además, es importante considerar el costo computacional asociado con la implementación de cada modelo. La NN, aunque precisa, requiere recursos significativos tanto en términos de hardware como de tiempo, lo que puede no ser factible en todas las instalaciones astronómicas. Por lo tanto, la elección del modelo debe balancear precisión, tiempo de entrenamiento y disponibilidad de recursos, según las necesidades específicas de la aplicación.

TABLE III
PRECISIÓN

Modelo	Precisión
Supported Vector Machine	95.36%
Neural Network	96.70%

A. Sub section de ejemplo

IV. CONCLUSIONES

En este trabajo, hemos demostrado la eficacia de los modelos de IA, específicamente las SVM y las NN, para la clasificación precisa de cuerpos celestes. Aunque ambos modelos muestran un alto rendimiento en la tarea de clasificación, la SVM destaca por su notable eficiencia computacional, siendo significativamente más rápida en el proceso de entrenamiento y prueba, y requiriendo menos recursos en comparación con la NN. Estos hallazgos subrayan que, a pesar de la ligera ventaja en precisión de la NN, la SVM puede ser la opción preferida en aplicaciones prácticas donde la eficiencia y la rapidez son cruciales. Así, la SVM emerge como una herramienta particularmente atractiva para la automatización de la clasificación de objetos celestes en grandes encuestas astronómicas.

REFERENCES

- [1] Bailer-Jones, C. A. L. (2011). *Stars and Their Spectra: An Introduction to the Spectral Sequence*. Cambridge University Press.
- [2] Zhang, Y., Zhao, Y., Zhao, G. (2009). "Automatic Classification of Stellar Spectra with Probabilistic Neural Network." *Publications of the Astronomical Society of the Pacific*, 121(882), 213-219.
- [3] York, Donald G., et al. 2000. "The Sloan Digital Sky Survey: Technical Summary." *The Astronomical Journal* 120 (3): 1579-1617.
- [4] fedesoriano. (January 2022). Stellar Classification Dataset - SDSS17, Retrieved April 10th, 2024 from: <https://www.kaggle.com/fedesoriano/stellar-classification-dataset-sdss17>