

Una Investigación Cualitativa sobre la Clasificación de Objetos Astronómicos mediante Machine Learning: Enfoque en Eficiencia y Ahorro en la Astronomía de Macrodatos

María de Jesús Cárdenas Chávez¹, Cynthia Alejandra Martínez Pinto^{2*}, Rosa María Michel Nava³, Ruth Clementina Barragán López⁴, Orestes Javier Pérez Cruz⁵. Tecnológico Nacional de México/Instituto Tecnológico de Cd. Guzmán: Depto. Ciencias Económico Administrativas¹, Depto. Sistemas y Computación^{2,3,4}, Estudiante de maestría en Ciencias y Computación⁵.

Resumen - Los proyectos de investigación deben sustentarse en una planificación estratégica y un riguroso análisis financiero con el propósito de mitigar posibles riesgos. El clasificador de objetos astronómicos está basado en algoritmos de inteligencia artificial de Machine Learning, utilizando datos de catálogo DR3 de la misión espacial Gaia, con la finalidad de generar ahorros en las operaciones de los proyectos que emplean estos métodos. En este tipo de proyectos, la inversión es significativa y es un factor necesario para el crecimiento económico. Sin embargo, es crucial que esta inversión se realice con el objetivo de ser rentable tanto en el aspecto económico como en el social.

Índice de Términos - Ahorro, inversión, investigación, clasificador de objetos astronómicos.

Abstract - Research projects should be grounded in strategic planning and thorough financial analysis to mitigate potential risks. The astronomical object classifier is based on artificial intelligence algorithms of Machine Learning, utilizing data from the DR3 catalog of the Gaia

space mission, with the aim of generating savings in project operations that employ these methods. In such projects, the investment is substantial and essential for economic growth. However, it is crucial that this investment is made with the goal of being profitable both economically and socially.

Keywords - Savings, investment, research, classifier of astronomical objects.

I. INTRODUCCIÓN

El rápido crecimiento en el tamaño y complejidad de los datos astronómicos recopilados por misiones espaciales y los telescopios terrestres ha introducido a la Astronomía en la era de la ciencia de los macrodatos. La inteligencia artificial es una de las herramientas más utilizadas en este campo del conocimiento (Martínez & Bertha, 2017), debido a la efectividad demostrada en el manejo de grandes volúmenes de datos de forma automatizada. Además, los avances tecnológicos están permitiendo al ser humano obtener grandes cantidades de información en

M.F. María de Jesús Cárdenas Chávez, docente del Departamento de Ciencias Económico Administrativas en el Tecnológico Nacional de México / Instituto Tecnológico de Cd. Guzmán, México. maria.cc@cdguzman.tecnm.mx (autor correspondiente).

Cynthia Alejandra Martínez Pinto, jefa de la división de estudios de postgrado en el Tecnológico Nacional de México / Instituto Tecnológico de Cd. Guzmán, México. cynthia.mp@cdguzman.tecnm.mx

Rosa María Michel Nava docente del Departamento de Sistemas y Computación en el Tecnológico Nacional de México / Instituto Tecnológico de Cd. Guzmán, México. rosa.cc@cdguzman.tecnm.mx

Ruth Clementina Barragán López, jefa de laboratorio de cómputo en el Tecnológico Nacional de México / Instituto Tecnológico de Cd. Guzmán, México. Ruth.bl@cdguzman.tecnm.mx

Orestes Javier Pérez Cruz, alumno de Maestría en Ciencias de la Computación en el Tecnológico Nacional de México / Instituto Tecnológico de Cd. Guzmán, México.

M22290971@cdguzman.tecnm.mx

proporciones difíciles de procesar. IBM ha publicado que cada día se generan de distintas fuentes más de 2,5 trillones de bytes de información y esta información cada día se genera más, rápido, variada y sin formato específico (Diario ABC, 2014). Dicha información se origina de diversas fuentes, siendo la Astronomía, YouTube y Twitter los tres mayores generadores de datos a nivel Big Data, cada uno representando retos particulares que requieren abordajes distintos (Stephens et al., 2015), 2015).

El procesamiento de estos datos masivos y su clasificación se ha vuelto esencial para los estudios astronómicos. Esto se logra a través de trabajos automatizados que utilizan métodos de las ciencias computacionales, como la inteligencia artificial y la Minería de Datos. Estas técnicas adoptadas por el campo del Big Data aprovechan modelos que se han perfeccionado a lo largo de los años para procesar las grandes cantidades de datos generados en la astronomía.

En este contexto, los avances tecnológicos no solo demandan inversión en capital intelectual para desarrollar sistemas como clasificadores de objetos astronómicos, sino que también generan ahorros significativos para los usuarios. Estos ahorros se manifiestan en términos de tiempo, calidad, eficiencia y oportunidad en los estudios astronómicos. Los algoritmos de Machine Learning utilizados para clasificar objetos astronómicos peculiares a través de su espectro permiten encontrar estrellas en etapas cruciales de su ciclo de vida de manera automática. Estos avances tecnológicos no solo aportan al crecimiento económico, sino también a la optimización de los recursos en el campo de la astronomía. Esta investigación cualitativa menciona las actividades que se realizarían para realizar la Clasificación de Objetos Astronómicos mediante Machine Learning.

II. METODOLOGÍA

Existen objetos astronómicos que presentan características peculiares y atraen la atención de los astrónomos.

Los métodos convencionales actuales para la clasificación de objetos astronómicos en la detección de estrellas denominadas peculiares están limitados en alcance, dado que estos objetos se encuentran a años luz de la Tierra y presentan características que dificultan su detección.

Con el desarrollo de esta investigación se logra obtener una clasificación de objetos astronómicos utilizando la información obtenida de sus espectros, empleando un

enfoque diferente a los existentes. Este consiste en seleccionar el modelo de Machine Learning con mejor resultado y rendimiento, que reciba como entrada los datos espectrales del catálogo DR3 de Gaia.

Este modelo resulta de gran utilidad para los astrónomos, ya que se convierte en una valiosa herramienta en la identificación y clasificación de estrellas peculiares.

Los objetos astronómicos mencionados anteriormente presentan dificultades en su identificación debido a sus características únicas y su baja representación en comparación con otros cuerpos celestes. Un recurso fundamental a disposición de los astrónomos es la espectroscopia, una técnica que permite analizar las composiciones químicas, propiedades físicas y velocidades radiales de las fuentes astronómicas. A través de técnicas espectroscópicas, es posible examinar el espectro de radiación electromagnética, abarcando desde la luz visible y ultravioleta hasta rayos X, infrarrojos y ondas de radio, emitidos por estrellas y otros objetos celestes (Massey & Hanson, 2013).

La Inteligencia Artificial se ha convertido en un imperativo en los tiempos modernos, donde la cantidad de datos que se maneja crece exponencialmente. Los algoritmos de clasificación, resultado del aprendizaje automático, se han vuelto cada vez más populares entre los astrónomos debido a su confiabilidad, certeza y rapidez.

Las características del modelo influirán en la posterior predicción, lo que implica un uso y resultado preciso. La ingeniería de características es el proceso de determinar las características más apropiadas dados los datos, el modelo y la tarea. Se debe definir la estructura de sus datos y tener características relevantes, debido a que los resultados dependerán de una adecuada elección.

El estudio se centró en la clasificación de los siguientes objetos astronómicos: Estrellas Simbióticas, Nebulosas Planetarias y Gigantes Rojas.

Las actividades que se realizaron para desarrollar el modelo, se muestran en la Figura 1



Fig. 1. Desarrollo del modelo clasificador de objetos astronómicos

III. RESULTADOS

Análisis y adquisición de los datos: La base de datos que se utilizó en la investigación contiene los espectros de baja resolución de alrededor de 470 millones de fuentes. Dicha base de datos puede ser obtenida del último catálogo Gaia DR3, liberado el 13 de junio de 2022 y la cual es de dominio público.

Se debe definir los objetos que se quieren clasificar, los cuales comparten las características de presentar líneas de emisión y ser difíciles de detectar, como nebulosas planetarias, estrellas simbióticas y objetos de la región HII. También se debe precisar el subconjunto de campos de la base de datos que son útiles.

Preparación de los datos: A los datos en bruto hay que realizarle un tratamiento previo, de modo que sean adecuados para su posterior procesamiento y análisis. Es necesario hacer una limpieza y etiquetado de datos en bruto en un formato adecuado para algoritmos de machine learning (ML). La preparación de datos puede ocupar hasta un 80 % del tiempo dedicado a un proyecto de ML.

Ingeniería de Características: Las características del modelo influirán en la posterior predicción, lo que implica un uso y resultado preciso. La ingeniería de características es el proceso de determinar las características más apropiadas dados los datos, el modelo y la tarea. Se debe definir la estructura de sus datos y tener características relevantes, debido a que los resultados dependerán de una adecuada elección.

Selección del modelo: En esta etapa se evalúan los posibles modelos de Machine Learning a comparar, tomando como referencias los antecedentes de estos en trabajos previos, donde obtuvieron un buen grado de confiabilidad.

Entrenamiento, Pruebas y Evaluación de Resultados: Se realiza un proceso de entrenamiento utilizando los modelos de Machine Learning previamente seleccionados. Para evaluar la precisión de los algoritmos se utiliza el método de validación cruzada (cross-validation), donde se definen subconjuntos, uno de ellos se separa del proceso de entrenamiento para ser utilizado en la etapa de validación. De esta forma se podría detectar si hubo un sobreajuste y el modelo no logró generalizar. Para evaluar los resultados obtenidos se establecen métricas que definen la efectividad del algoritmo y su rendimiento. Este proceso es cíclico, y en cada iteración se van ajustando los parámetros hasta obtener un resultado admisible.

Los estudios de ahorro e inversión: al utilizar un clasificador de objetos astronómicos son vitales para proponer investigaciones de este tipo ya que la inversión es un factor necesario para el crecimiento de la economía, pero no será suficiente si no se hace de manera rentable, cumpliendo sus propósitos ya sea de rentabilidad económica o social mediante la creación de capacidades que posibiliten otras inversiones.

La planeación financiera, se nutre del análisis económico en el cual se realizan proyecciones de las diversas decisiones de inversión, financiamiento, presupuestación y resultados esperados.

Los principales recursos que se aplican a una inversión son el tiempo, el trabajo y el capital.

Por otro lado, la decisión de llevar adelante una inversión dependerá del análisis que realice el sujeto inversor acerca de 3 variables fundamentales: riesgo, rendimiento y tiempo (Enciclopedia Económica, s. f.).

La importancia del ahorro radica en que permite generar disciplina respecto al manejo eficiente del dinero, lo cual conlleva a una óptima organización de los ingresos.

De igual manera, el ahorro es una inversión a largo plazo que nos permite respaldarnos para el futuro, garantizando que podamos afrontar situaciones imprevistas o no programadas, así como también la realización de ciertos proyectos planeados (Enciclopedia Económica, s. f.).

El ahorro es una práctica usual y también un concepto importante en la teoría económica, comprendido como el porcentaje de la renta o del ingreso que no se destina al consumo. Es por lo que existen distintas formas de ahorro

e incluso se han diseñado instrumentos financieros cuyo rol específico es permitir o incrementar el ahorro deseado.

El ahorro es una actividad de planificación económica vital para la supervivencia de un sistema productivo en el tiempo, ya que entraña la posibilidad de que parte de los recursos producidos no sean consumidos o despilarrados, sino estratégicamente salvaguardados para el futuro.

Por eso el ahorro es fomentado a todo nivel, ya que entraña una utilización más sensata y previsor de los recursos disponibles, que sirva para enfrentar necesidades venideras o que pueda invertirse en nuevos proyectos («Ahorro - Concepto, tipos, importancia y qué es inversión», s. f.).

La inversión para llevar a cabo este tipo de proyectos en su mayoría es intangible ya que depende del capital intelectual que en el proyecto de investigación. Para realizar el clasificador de objetos astronómicos en el Instituto Tecnológico de Cd. Guzmán, ha sido necesario un doctorante en tecnologías de la información, un maestrante en ciencias de la computación, un tesista en Ingeniería en Sistemas Computacionales, 3 residentes en Ingeniería en Sistemas Computacionales. Esto significa una inversión en capital intelectual que se traduce financieramente en recursos económicos llega a ser significativa, sin embargo, la rentabilidad se mide principalmente en el impacto social que este proyecto genera.

Así mismo, el ahorro al usar el método propuesto se mide principalmente por el tiempo que utiliza el usuario para la clasificación de objetos astronómicos, la calidad en los resultados y por ende la eficiencia y oportunidad en el diagnóstico de sus trabajos.

Los ahorros anteriores finalmente impactan financieramente en beneficio del usuario del método ya que la diferencia al utilizar un clasificador tradicional y un clasificador a través de trabajos automatizados, con métodos provenientes de las ciencias computacionales es traducido en grandes beneficios financieros que cada usuario pudiera cuantificarlos, en general, el ahorro en el uso de la tecnología es muy significativo actualmente por los impactos económicos y sociales que este genera.

IV. CONCLUSIONES

El desarrollo de la investigación de un clasificador de objetos astronómicos se divide el trabajo en diferentes etapas: fase de adquisición y análisis de los datos, preparación de los datos, ingeniería de características, selección del modelo, entrenamiento, pruebas, mejora y evaluación de los resultados, los cuales dependen de una inversión de capital intelectual de diferentes especialistas los cuales representan significativamente un gran apoyo para el impacto social que genera el método sugerido para el uso de los astrónomos.

Existen estrellas un tanto peculiares, que debido a sus características son difíciles de clasificar con los métodos convencionales. Esto, sumado a la gran cantidad de información de espectros que un telescopio moderno puede generar, impide a los astrónomos procesar dichos datos de manera individual. La clasificación automática es un imperativo en los tiempos actuales, donde se maneja un gran volumen de datos.

Aplicar estrategias básicas en las organizaciones apoya al logro de grandes éxitos, por lo tanto, es necesario crear una cultura de proyecciones a corto y largo plazo para una excelente investigación que no se trunque por falta de inversión cualitativa y cuantitativa con el fin de que genere las condiciones de ahorro óptimo para los usuarios del método.

Usar tecnología para el procesamiento y análisis de datos impacta significativamente en el desarrollo económico, ecológico y social.

RECONOCIMIENTO

Agradecemos al Tecnológico Nacional de México por el apoyo recibido en el proyecto financiado “Algoritmos de Inteligencia Artificial orientados a la clasificación de espectros astronómicos”, con número “17243”, además de las directrices que orientan nuestros esfuerzos en bien del país y los jóvenes estudiantes. También expresamos nuestra sincera gratitud a las honorables autoridades del Instituto Tecnológico de Ciudad Guzmán, incluyendo al Director, M.C. Sergio Octavio Rosales Aguayo, y a nuestra Subdirectora Académica, Dra. María Guadalupe Sánchez Cervantes.

REFERENCES

- [1] C. Martínez y A. Bertha, «Detección de objetos variables en imágenes astronómicas a través del tiempo», jun. 2017, Accedido: 4 de octubre de 2023. [En

línea]. Disponible en:
<http://tesis.ipn.mx/xmlui/handle/123456789/22204>

[2] Ep. (2014, 23 abril). Cada día se generan más de 2,5 trillones de bytes de información. Diario ABC. <https://www.abc.es/tecnologia/redes/20140423/abci-trillones-byte-informacion-cada-201404222207.html> [3]

Z. D. Stephens et al., «Big Data: Astronomical or Genomical?», PLOS Biol., vol. 13, n.o 7, p. e1002195, jul. 2015, doi: 10.1371/journal.pbio.1002195.

[4] P. Massey y M. M. Hanson, «Astronomical Spectroscopy», en *Planets, Stars and Stellar Systems: Volume 2: Astronomical Techniques, Software, and Data*, T. D. Oswalt y H. E. Bond, Eds., Dordrecht: Springer Netherlands, 2013, pp. 35-98. doi: 10.1007/978-94-007-5618-2_2.

[5] «Enciclopedia Económica», Enciclopedia Económica. Accedido: 4 de octubre de 2023. [En línea]. Disponible en: <https://enciclopediaeconomica.com/>

[6] «Ahorro - Concepto, tipos, importancia y qué es inversión», <https://concepto.de/>. Accedido: 4 de octubre de 2023. [En línea]. Disponible en: <https://concepto.de/ahorro/>