



UNIVERSIDAD DE GRANADA

Facultad de Ciencias

GRADO EN FÍSICA

TRABAJO FIN DE GRADO

**Procesado de imágenes del prototipo del
Large-Sized-Telescope del
Cherenkov Telescope Array Observatory
mediante algoritmo de hormigas**

Presentado por:

D. Germán Cano Amaro

Tutor:

D. Alberto Guillén Perales

Co-tutor:

D. Rubén López Coto

Curso Académico 2023/2024

Resumen

Los telescopios de imagen atmosférica Cherenkov son una de las principales herramientas para el estudio de los rayos gamma de altas energías, cruciales para el estudio de numerosos fenómenos y objetos en el universo. Bajo este contexto, el Cherenkov Telescope Array Observatory (CTAO), planea construir 64 de estos dispositivos repartidos entre ambos hemisferios terrestres. La sensibilidad requerida para poder observar la radiación Cherenkov conlleva el registro de grandes cantidades de datos en cada observación. Con el objetivo de reducir el coste de almacenamiento de dichos datos se muestra en este estudio la posibilidad de expresarlos de forma alternativa mediante grafos. Estos grafos se procesaron posteriormente con el algoritmo de las hormigas para encontrar su representación óptima. Dicha representación redujo el espacio de almacenamiento en memoria aproximadamente un 50% respecto a otros formatos comúnmente usados. Puesto que para su posterior estudio los datos requieren una previa clasificación se propone el uso de distintos algoritmos de machine learning para este fin. La capacidad de los modelos de clasificar los datos apenas varía entre los datos originales y su representación mediante grafos, es por ello que se proponen para futuros estudios tanto distintas reinterpretaciones de los grafos como distinta estructura de los propios modelos. El código utilizado para generar los resultados aquí mostrados se puede encontrar en: <https://github.com/german-cano/CTAO-image-processing-with-ants-algorithm.git>

1 Introducción

El universo, sus objetos y fenómenos han sido y son materia de estudio constante. Entenderlos nos ayuda a comprender nuestro entorno con mayor perspectiva tanto pasada como futura. Para poder realizar las observaciones necesarias desde la tierra que nos permitan obtener información del universo se suele recurrir a las señales electromagnéticas que de él nos llegan. De este modo, según qué queramos estudiar podremos fijarnos en una parte del espectro u otra. Uno de los motivos principales para recurrir al espectro electromagnético es, a parte de la información que aporta, su capacidad para viajar por el espacio desde la fuente hasta nosotros sin apenas desviarse. Dentro de este espectro tienen una importante relevancia los rayos gamma de altas energías pues, al contrario de las partículas cargadas que puedan llegar a la atmósfera terrestre, los rayos gamma no se ven desviados por los campos magnéticos que encuentren desde su origen hasta la tierra. Los rayos gamma de alta energía se emiten en los procesos más violentos del universo por objetos, tanto galácticos como extragalácticos, como supernovas, púlsares o estrellas de neutrones entre otros. Estudiar estos fenómenos conlleva estudiar la formación de estrellas, las interacciones entre estas o entre ellas y otros objetos como agujeros negros, así como una mayor comprensión de la composición del universo. Los rayos gamma de altas energías se pueden medir mediante detectores en el espacio orbitando la tierra o mediante detectores a nivel terrestre. Si bien los primeros son eficientes y proporcionan resultados satisfactorios, el precio que acarrearán hace que la mayoría de detectores se encuentren a nivel terrestre. No obstante, las mediciones con estos implican, debido a la alta energía de los rayos gamma de interés, que no podamos medirlos directamente. Al entrar en la atmósfera, estos rayos interactúan con ella, debido a su alta energía generan pares electrón-positrón que a su vez dan lugar a más rayos gamma que a base de repetir este proceso dan lugar a una cascada (Bremsstrahlung) que hace imposible medir directamente el rayo gamma originario. Los detectores deben buscar por tanto formas indirectas de obtener información del evento, en concreto suelen aprovechar que, debido a la alta energía de los rayos, los pares electrón-positrón producidos serán también suficientemente energéticos como para emitir radiación Cherenkov. Esta radiación es lo que realmente detectan los sistemas que debido a este proceso se conocen como IACTs, por sus siglas en inglés Image Atmospheric Cherenkov Telescopes.

A lo largo de las últimas décadas, a medida que la tecnología de los IACTs se iba desarrollando, varias matrices de telescopios (MAGIC, VERITAS y H.E.S.S.) han estado produciendo los resultados más importantes en el campo. Sus resultados han mostrado la importancia de seguir investigando en este campo y han servido de base para la creación de CTAO (Cherenkov Telescope Array Observatory) [1]. CTAO es el resultado de la unión de instituciones de 25 países para construir el mayor observatorio de IACTs con equipos de detección en ambos hemisferios terrestres. El primer prototipo de estos telescopios se encuentra ya instalado y operativo en el observatorio de El Roque de los Muchachos, La Palma. Por otro lado, los IACTs se enfrentan a un problema que dadas las dimensiones de CTAO se hará cada vez más acuciante y requiere de una solución. Para poder

detectar la radiación Cherenkov los IACTs deben ser muy sensibles a cualquier señal que llegue a sus sensores, esto provoca que junto a los rayos gamma de alta energía que queremos estudiar capten cualquier otra señal, es decir, el volumen de ruido captado en las mediciones es tal que supone un gran problema para el posterior tratamiento, clasificación y almacenamiento de los datos. Actualmente el prototipo de telescopio en funcionamiento de CTAO registra en torno a 2 TiB/h pudiendo llegar a los 25 TiB en una noche. Solo para este tipo de prototipo se espera cuadruplicar estas cifras cuando estén todos los telescopios en funcionamiento. Es por esto que es necesario encontrar una forma de registrar los datos que nos permita, sin perder información, reducir el coste de su procesado y almacenamiento, así como el desarrollo de técnicas y algoritmos que faciliten la clasificación automática de los datos registrados para su posterior estudio.

Se propone en este estudio la aplicación del algoritmo de las hormigas como medio para obtener una representación alternativa de los datos a través de grafos, cabe mencionar que hasta el momento de la redacción de este trabajo no se han encontrado otros estudios en los que se explore posibilidad de representar los datos obtenidos por IACTs mediante grafos. Los datos utilizados provienen del prototipo ya en uso de La Palma y se comparará el espacio en memoria de estos con la correspondiente representación por grafos. Así mismo se estudiará la implementación de varios algoritmos de machine learning para la clasificación tanto de los datos originales como de la obtenida mediante el algoritmo de las hormigas.

2 Objetivos

Se desglosan a continuación, por claridad en su exposición, los objetivos principales del presente estudio:

- Exponer de forma breve las principales fuentes de rayos gamma de alta energía.
- Explorar los sistemas de detección actuales de rayos gamma de alta energía, IACT, con especial atención en CTAO y explicar el proceso de toma de datos.
- Analizar y tratar el conjunto de datos reales del prototipo de telescopio ya en funcionamiento de CTAO.
- Plantear y obtener diversas representaciones alternativas mediante el uso de grafos del conjunto de datos.
- Comparar la diferencia de espacio en memoria de la representación de los datos antes y después de su representación mediante grafos.
- Clasificar, mediante algoritmos de machine learning, ambos conjuntos de datos en función del tipo de evento al que pertenecen.

3 Física de astropartículas

Para poder tener una idea general del origen de las observaciones realizadas por CTAO necesitamos comprender algunos aspectos básicos de la física de astropartículas. De otro modo sería imposible realizar tanto una correcta clasificación de dichos datos como su posterior estudio para determinar el origen del evento del que provienen.

Dichas observaciones provienen del 5 % del universo, que corresponde a la materia bariónica que forma la parte del universo que conocemos. El resto de componentes del universo se dividen en energía oscura 69 % y materia oscura 26 %, las cuales sabemos que existen por sus contribuciones a la expansión del universo y por sus efectos gravitatorios respectivamente. A día de hoy no existe una respuesta clara y consensuada sobre qué forma este 95 % del universo desconocido aunque algunos autores como [2], [3], [4], [5] proponen agujeros negros primitivos, neutrinos estériles o hexaquarks como posibles candidatos entre otros. Por su parte, el universo conocido lo podemos clasificar en universo térmico y no térmico. El universo térmico lo componen las estrellas, el polvo y gas estelar, la radiación térmica, el fondo cósmico de microondas, etc. El no térmico, en cambio, lo forma la radiación generada en procesos no térmicos como interacciones entre partículas o de partículas con campos magnéticos. Estas radiaciones suelen tener energías superiores a la que podría alcanzar cualquier radiación térmica, del orden de los ZeV y componen una parte importante de la energía del universo conocido. Pueden surgir de eventos como las colisiones entre agujeros negros masivos con estrellas, entre estrellas de neutrones, entre dos agujeros negros, supernovas,... La física de astropartículas se encarga del estudio de la interacción entre partículas cosmológicas de muy alta energía cuya comprensión es fundamental para entender la formación y evolución de las galaxias. Algunas de las formas de esta radiación son los rayos cósmicos (de origen aún desconocido), los fotones (producidos en decaimientos, scattering, radiación sincrotrón, ...), neutrinos y ondas gravitacionales.

3.1 Rayos cósmicos

Los rayos cósmicos son partículas cargadas que llegan a nuestra atmósfera de manera casi isotropa. La mayoría de ellos son protones, entorno al 90 % y núcleos de Helio, 9 %, el 1 % restante está compuesto de núcleos más pesados, electrones, positrones... . Debido a su carga, sufren desviaciones por los distintos campos magnéticos que encuentran a su paso en el trayecto desde su origen hasta nosotros. Esto hace imposible trazar su trayectoria y localizar dicho origen. El descubrimiento de los rayos cósmicos en 1912 le valió, en 1936, el premio Nobel de Física a Victor Hess [6]. Este, mediante mediciones de la ionización atmosférica en globos aerostáticos observó cómo la ionización aumentaba con la altura, su conclusión fue que esta se debía a radiación extraterrestre.

El consecuente estudio de los rayos cósmicos llevó al descubrimiento de partículas desconocidas hasta el momento como fue el caso del positrón (1932) [7],

el muón (1937) [8], el pión (1947) [9], el kaón (1947) [10] o el barión Λ (1951) [11]. Los rayos cósmicos eran el único modo de estudiar dichas partículas hasta que a mitad de siglo se construyeron los primeros aceleradores de partículas. No obstante, incluso a día de hoy, la física de astropartículas juega un papel fundamental en el estudio de estas partículas pues la máxima energía que somos capaces de estudiar en los aceleradores actuales es del orden de los TeV. En contraposición, los rayos cósmicos pueden alcanzar energías mayores llegando a los PeV para los rayos extragalácticos. Sin embargo, desconocemos el mecanismo por el cual estos rayos alcanzan tales energías. Las ondas de choque creadas en las supernovas han sido durante mucho tiempo uno de los principales candidatos como acelerador los rayos cósmicos [12], [13], aunque a día de hoy no se ha podido demostrar aún. Además, una vez emitido, la interacción constante con los distintos campos magnéticos de la galaxia puede modificar sensiblemente la energía del rayo, acelerándolo o frenándolo.

Aunque no podamos determinar la fuente exacta de los rayos, sabemos que, por ejemplo, los electrones primarios son producidos en parte en remanentes de supernovas mientras que los secundarios son producidos en las colisiones de los rayos cósmicos con la materia. Por su parte, los positrones de energías inferiores a los GeV también son producidos en estas colisiones mientras que aquellos con energías superiores se observa que deben provenir de una fuente primaria, algunos autores proponen que esta fuente primaria sea materia oscura [14], [15]. Otra cuestión abierta sobre los rayos cósmicos es la rigidez (cociente entre energía y carga) de especies primarias como el oxígeno o el carbono o secundarias como el Boro, Litio y Berilio. Estos ejemplos ponen de relieve la importancia del estudio de los rayos cósmicos y su origen para lo cual tendremos que usar por tanto partículas neutras como neutrinos, ondas gravitacionales o rayos gamma.

3.2 Rayos Gamma

Los fotones producidos en procesos no térmicos no tienen por qué tener siempre altas energías, se producen también mediante aceleración sincrotrón (partículas aceleradas en un campo magnético) fotones en los espectros de radio, infrarrojo y rayos X. Los fotones de altas energías, rayos gamma, se producen principalmente en scattering Compton inverso y decaimiento del π^0 .

3.3 Fuentes galácticas

La formación estelar parte de una nebulosa cuyo material se va agrupando hasta formar una estrella. En el caso de que la nebulosa tenga suficiente material la estrella será masiva y su vida terminará en una supernova dando lugar a una estrella de neutrones o un agujero negro para el caso de estrellas supermasivas. Si la nebulosa no parte de tanto material la estrella tendrá una masa media-baja y morirá en una nebulosa planetaria dando lugar a una enana blanca.

3.3.1 Remanentes de Supernovas

Se distinguen principalmente tres tipos de remanentes de supernova, aquellos en los que se puede observar la onda de choque de la supernova, los que forman una nebulosa a partir de un púlsar y la combinación de ambas.

3.3.2 Estrellas de neutrones, púlsares

Tras la supernova de una estrella masiva queda una estrella de neutrones, estas son estrellas altamente magnetizadas de unos 10 km de radio que rotan sobre su eje de rotación rápidamente, varias veces por segundo. Estas estrellas se conocen como púlsares cuando su principal fuente de emisión es su rotación. Esta emisión puede darse desde frecuencias de radio a rayos gamma de alta energía. Uno de los púlsares más estudiado por la comunidad es la nebulosa del cangrejo la cual se suele usar como calibrador de rayos gamma de alta energía. Desde que se comenzó a estudiar la nebulosa del cangrejo se rompió la barrera de la que se pensaba que era la energía máxima a la que podía emitir. Así, en 2008 MAGIC midió emisiones por encima de los 25 GeV [16], en 2011 VERITAS midió emisiones por encima de los 100 GeV [17], en 2012 MAGIC llegó a medir hasta 400 GeV [18] y nuevamente en 2016 MAGIC midió energías del orden de los TeV [19]. A día de hoy no hay explicación para este rango de emisiones. Tras la nebulosa del cangrejo se observaron otros púlsares que emitían rayos gamma de altas energías como Vela o Geminga. El caso de Geminga fue inesperado pues es el primer pulsar de media edad (~ 300 kyr) que se observa emitir en altas energías, entre 15 y 80 GeV [20].

Si bien desconocemos cómo estos rayos gamma alcanzan estas altas energías en los púlsares, conocemos cómo se producen dichos rayos. En su rotación, la estrella de neutrones genera un viento de electrones y positrones que debido a los fuertes campos magnéticos y a la rotación llegan hasta el material remanente de la supernova interactuando con él y emitiendo los rayos gamma mediante radiación sincrotrón y scattering compton inverso.

3.3.3 Sistemas binarios y otras fuentes

Otra de las fuentes rayos gamma de alta energía son los sistemas binarios, compuestos por una estrella de neutrones o un agujero negro junto a otra estrella de neutrones u otro agujero negro. Para el caso de sistemas de agujeros negros junto a estrellas de neutrones, cuando el agujero negro atrae material de su compañera, emite haces de rayos gamma que podemos medir. Así mismo, para agujeros negros que no forman parte de sistemas binarios como el agujero negro supermasivo en el centro de la vía láctea, podemos estudiar la emisión de rayos gamma debida a la aceleración de partículas por el agujero negro. Se han medido rayos gamma de energías superiores a los 40 TeV. Es posible encontrar una explicación más detallada de cada fuente aquí tratada entre otras en la la H.E.S.S galactic plane survey [21].

3.4 Fuentes extragalácticas

Debido a su lejanía, estudiar las fuentes de rayos gamma extragalácticas supone una mayor complejidad con respecto a las galácticas. Para poder observar estas señales por tanto es necesario que la fuente, la cual se presupone que es el agujero negro supermasivo en el centro de otra galaxia, tenga una gran actividad y emita por tanto rayos gamma suficientes y con suficiente energía como para que podamos detectarlos. Una de las causas para la pérdida de información proveniente de estas fuentes es la interacción de los rayos gamma con fotones del fondo cósmico creando pares electrón-positrón. Esto, hace que gran parte de los rayos gamma extragalácticos que observamos sean en realidad productos menos energéticos del rayo original. Otra de las fuentes extragalácticas que podemos observar son las explosiones de rayos gamma (Gamma-ray Burst) causadas por la fusión de un sistema binario de estrellas de neutrones o por el colapso de una estrella supermasiva en un agujero negro.

4 Técnicas de detección: IACT

Actualmente se usan dos tipos de telescopios para observar los rayos gamma de alta energía. Los telescopios de imagen atmosférica por radiación Cherenkov (IACTs) y los detectores de superficie o volumen (Detectores de Partículas). Mientras que los primeros son direccionales, con gran resolución angular ($<5'$) y campo de visión pequeño (3° - 5°) los segundos tienen menor resolución angular ($>10'$) y su campo de visión cubre el cielo en su totalidad. Además, los surveys tienen una monitorización del cielo constante mientras que los IACTs solo pueden monitorear durante la noche en buenas condiciones de visibilidad. En contraposición los IACTs ofrecen una mayor precisión y resolución debido a su gran sensibilidad. No obstante algunos sistemas survey muestran buenos resultados en la detección de rayos gamma, en concreto los satélites que funcionan como calorímetros. Sin embargo, el coste de poner un satélite de estas características en órbita restringe enormemente el tamaño de este y por ende su ratio de detección y su utilidad. Análogamente, los sistemas de detección a nivel terrestre usan el mismo principio que los calorímetros bajo la salvedad de que en este caso es la propia atmósfera la que actúa como calorímetro. El sistema detecta por tanto la cascada generada por el rayo gamma al entrar en contacto con la atmósfera y, a partir de ella, reconstruye el rayo gamma original. Estos sistemas se aprovechan, por tanto, de que la atmósfera no es lo suficientemente gruesa como para que la cascada no sea detectada a nivel terrestre.

Cuando un rayo gamma entra en la atmósfera, interactúa con ella creando un par electrón-positrón, cada uno de estos productos emite sucesivamente otro fotón por bremsstrahlung el cual, si es lo suficientemente energético, volverá a crear otro par electrón-positrón. Se crea así una cascada de interacciones que continúa hasta que la energía perdida por la ionización de los pares electrón-positrón supera a la energía de los fotones. Si lo que llega a la atmósfera es, en cambio, un rayo cósmico, este interactuará con algún núcleo atmosférico desencadenando una

cascada de partículas distintas. Dichas partículas interactuarán a su vez con otros núcleos repitiéndose el proceso. Es importante notar que en una cascada hadrónica, al ser los productos de las colisiones distintos, también lo serán la forma de la cascada de partículas que estos produzcan. El tiempo que tardará cada subproducto en llegar a tierra será, por tanto, distinto. Independientemente de si el rayo que entra en la atmósfera es cósmico o gamma, debido a su alta energía los productos de ambas cascadas viajarán a una velocidad mayor que la de la luz en el aire emitiendo por tanto radiación Cherenkov. Los IACTs detectan esta radiación de los productos secundarios a partir de la cual reconstruyen el rayo original.

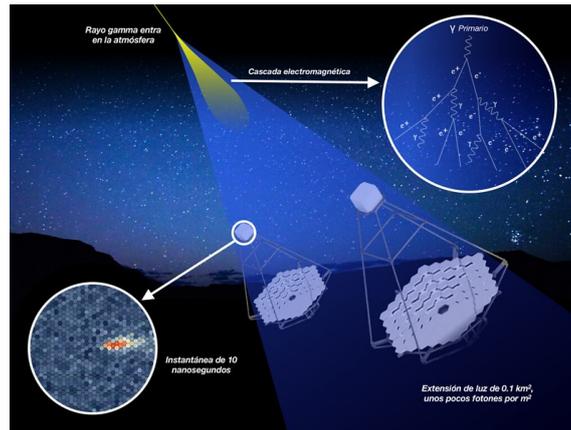


Figura 1: Esquema gráfico de la medida por un sistema IACT de radiación Cherenkov causada por un rayo gamma

Para un rayo gamma de energía aproximadamente 1 TeV llegan al nivel terrestre en una ráfaga de pocos nanosegundos una densidad aproximada de fotones de $100 \text{ m}^{-2} \text{ s}^{-1}$ es por esto que los IACTs deben tener gran sensibilidad. Para ello, cuentan con un gran reflector el cual les permite captar la mayor cantidad posible de fotones mediante una cámara hecha de fotodetectores, los cuales son capaces de detectar hasta un 50 % de los fotones únicos que llegan a ellos, y un sistema de accionamiento y adquisición de datos veloz. En concreto, los fotodetectores que usan la mayoría de IACTs son fotomultiplicadores de tubo, estos, transforman los fotones que captan en una cascada de electrones que podemos medir mediante ciertos dispositivos electrónicos. Para esto se sirven de un fotocátodo el cual a través del efecto fotoeléctrico emite electrones gracias a los fotones captados, no obstante estos electrones son insuficientes para poder medirlos de forma que se hace uso de dínodos, metales que al recibir electrones acelerados emiten mayor cantidad de estos. De esta forma redirigiendo los electrones por sucesivos dínodos se obtiene una señal eléctrica medible a partir de unos pocos fotones [22]. A parte de los fotodetectores, los IACTs están dotados de un amplio conjunto de subsistemas que permiten al telescopio actuar con rapidez o trabajar bajo condiciones meteorológicas adversas. No obstante, uno de los sistemas de mayor importancia son los algoritmos de análisis de la señal que nos permiten, a partir de la señal recogida del fotodetector, reconstruir el evento. Esto, incluye algoritmos de procesamiento de la señal y de limpieza de los datos. Mediante el procesamiento de la señal se busca obtener el número de fotoelectrones y el tiempo de llegada por cada pí-

xel del sensor durante la medición de un evento. El tiempo de medición de los eventos suele ser de algunas decenas de nanosegundos puesto que se busca medir el mayor número de fotones posibles del evento sin medir ruido en exceso. Para conseguir esto, el algoritmo trata de encontrar el mejor ajuste entre la señal recogida por cada píxel y una cierta función respuesta.

$$G = \sum_{i \in n} (S_i - NR_{i,T})^2 \quad (4.1)$$

Siendo G la bondad de ajuste entre la señal S_i y la función respuesta $R_{i,T}$ normalizada por un factor N con n el conjunto de datos y T un tiempo de medida concreto. Según el interés del estudio la función respuesta puede tomar formas diversas pero normalmente siempre se trata de ajustar N y T puesto que son los parámetros que nos dan un tiempo y una amplitud de la señal. El segundo algoritmo de gran importancia para los IACTs es el que se encarga de la limpieza de los píxeles, es decir, de determinar que píxeles tienen una carga de fotoelectrones provenientes de un evento y cuales tienen una carga debida al ruido. Nuevamente, hay varias opciones para implementar este algoritmo, una de las más básicas consiste en aproximar el tiempo de llegada de los fotones a una expresión lineal, este método se conoce como parametrización de Hillas [23]. Otros algoritmos seleccionan un píxel cuya carga de fotoelectrones corresponda de manera certera a un evento y estudian la carga de los píxeles colindantes bajo ciertas restricciones sobre su carga y la de sus vecinos para determinar si también pertenecen a dicho evento. Tras la limpieza de los datos se determina si el evento corresponde a un rayo gamma o un rayo cósmico, para lo cual nuevamente, se suele recurrir a diversas metodologías desde la clasificación mediante algoritmos de Machine Learning [24], [25], uso de métodos Monte-Carlo, extensiones de la parametrización de Hillas [26] o incluso por el ajuste de la representación de la carga registrada sobre cada píxel al la sección transversal del cono de emisión de la radiación. Por último, una vez se determina que el evento observado corresponde a un rayo gamma se estudia la energía del rayo primario y su ángulo de incidencia. El cono de emisión en el que está contenida la radiación Cherenkov de un rayo gamma que entra en la atmósfera tiene un diámetro aproximado de 300 m, esto hace que el área efectiva de cada IACT sea la correspondiente a un radio de 150 m entorno al telescopio. La imagen obtenida por el IACT será, por tanto, la sección transversal de dicho cono que, salvo que el telescopio este justo en el eje central de este, se corresponderá con una elipse. Es por esto que, por lo general, suelen situarse varios IACTs de modo que puedan captar el mismo evento desde perspectivas distintas para una mejor reconstrucción estereográfica. A partir de esta reconstrucción es posible determinar también el ángulo de incidencia del rayo, los sistemas actuales son capaces de obtener una precisión inferior a 0.1 para rayos superiores al centenar de GeV teniendo en cuenta los posibles errores instrumentales y estadísticos. Por su parte en los rayos cósmicos, debido a la la distinta energía de los subproductos de la cascada hadrónica, la radiación Cherenkov llega los IACTs de forma no homogénea, en tiempos distintos, perdiéndose así la característica sección elipsoidal de las observaciones de rayos gamma [27].

4.1 Sistemas actuales de IACTs

Tras el descubrimiento de la radiación Cherenkov a mediados del siglo pasado se tardó entorno a una treintena de años en construir el primer IACT (1989). Los IACTs han ido desarrollándose desde entonces hasta el día de hoy donde ya está en marcha la implementación de la cuarta generación de estos sistemas. Mientras tanto, la tercera generación que sigue en activo, está compuesta por tres proyectos esparcidos por el globo.

4.1.1 H.E.S.S.

Situado en Namibia lo componen cuatro IACTs de 12 m de diámetro más un quinto de 28 m. Opera desde el año 2004 estudiando señales desde los 30 GeV hasta los 100 TeV. El quinto telescopio, por su tamaño, detecta muy bien los rangos energéticos más bajos. Sin embargo, al haber solo uno es difícil reconstruir las características del evento.

4.1.2 VERITAS

En activo desde 2007, este observatorio de Arizona está compuesto por cuatro IACTs de 12 m de diámetro. Estudia un rango de energías que va desde los 85 GeV hasta algo más de los 30 TeV. Está expuesto a un 1 % de la nebulosa del cangrejo cada 25 horas.

4.1.3 MAGIC

MAGIC es una colaboración de 170 investigadores de 10 países distintos, situado en el observatorio de El Roque de los Muchachos en las Palmas de Gran Canaria, España. Lo componen dos IACTs de 17 m de diámetro ultraligeros a modo de poder redireccionarse velozmente para observar los distintos eventos. Gracias a esta capacidad de orientación en 2021 MAGIC observó una explosión de rayos gamma [28]. Este sistema de IACTs opera a partir de los 50 GeV hasta los 10 TeV con una resolución energética de entre 15 %-23 % y está expuesto a la nebulosa del cangrejo un 0.6 % cada 50 horas.

4.2 Cherenkov Telescope Array observatory, CTAO

CTAO es un esfuerzo global con el propósito de construir la próxima generación de IACTs. Hay más de 1500 científicos de 25 países involucrados incluyendo a los equipos de HESS, VERITAS y MAGIC. Este proyecto busca construir 64 telescopios entre el hemisferio sur, en el desierto de Atacama, Chile y en el hemisferio norte en la misma localización actual de MAGIC. Estos telescopios tendrán distintos tamaños para poder cubrir con mayor precisión un mayor rango de energías, así, se dividirán en cuatro LST (Large Size Telescope) de 23 m de diámetro situados en La Palma para cubrir el rango de bajas energías 20 GeV a 150 GeV, 23 MST

(Middle Size Telescope) de 12 m de diámetro distribuidos entre ambos hemisferios cubriendo el rango 150 GeV a 5 TeV y 37 SST (Small Size Telescope) de 4 m en el hemisferio sur para rangos superiores a los 5 TeV. CTAO será por tanto un observatorio abierto y no un proyecto de investigación, es decir, el observatorio de CTAO tendrá un tiempo limitado para realizar observaciones privadas tras lo cual las observaciones realizadas serán las propuestas por los distintos equipos científicos que las soliciten. El primer telescopio LST de este proyecto ya está operando en La Palma, con un área de espejo de 400 m² es capaz de detectar eventos a partir de los 20 GeV [29]. Al igual que MAGIC estos sistemas son muy ligeros.

Con este observatorio se espera ahondar nuestro conocimiento en los mecanismos de aceleración de partículas llegando a medir energías del orden de los PeV, buscando determinar el rol de dichas partículas en la formación de estrellas o la evolución de galaxias. Así mismo, se pretende estudiar con mayor detalle las estrellas de neutrones, los agujeros negros o el vacío cósmico, intentando dar explicación a problemas aún abiertos de la física.

5 Algoritmo de las hormigas, ACS

En ciertas especies de animales encontrar el camino más corto entre dos puntos puede suponer un gran valor adaptativo pues esto conlleva un ahorro de tiempo, esfuerzo y exposición a posibles predadores. Algunas especies que, como las hormigas, se mueven bajo tierra, recurrentemente tienen que decidir entre dos o más posibles caminos a tomar. Esto, también ocurre si encuentran un obstáculo como una piedra en su camino a recolectar comida para el nido. En el caso de las hormigas, dejan rastros de feromonas para solventar esta problemática. Para entender mejor este fenómeno supongamos el ejemplo de las hormigas que deben rodear un obstáculo para llegar a su destino. Al llegar a dicho obstáculo cada hormiga podrá rodearlo por la izquierda o por la derecha, supongamos que el camino al rodearlo por la derecha es más corto que al hacerlo por la izquierda. Las primeras hormigas que lleguen al obstáculo elegirán al azar uno de los dos caminos, a lo largo de su recorrido irán depositando feromonas. Según pasan hormigas, al ser el camino derecho más corto, las hormigas que lleguen a la bifurcación sentirán con mayor intensidad las feromonas de este camino, lo que hará que se decanten por recorrerlo añadiendo a su vez más feromonas e influenciando así al resto de hormigas. En última instancia todas las hormigas rodearán el obstáculo por la derecha habiendo encontrado así el camino más corto.

En la actualidad nos enfrentamos constantemente a problemas de optimización similares, en concreto el problema del vendedor ambulante (TSP) es de especial interés en este caso pues fue el primero sobre el que se aplicó el algoritmo de las hormigas. El problema del vendedor ambulante consiste en encontrar el recorrido óptimo entre un conjunto de ciudades que un vendedor ambulante puede recorrer sin pasar dos veces por la misma ciudad. El algoritmo de las hormigas [30], [31] se basa en el comportamiento de las hormigas reales para encontrar el camino más corto entre dos puntos. Para ello se genera un conjunto de agentes (hormigas) las cuales trabajaran en paralelo para encontrar de forma heurística la

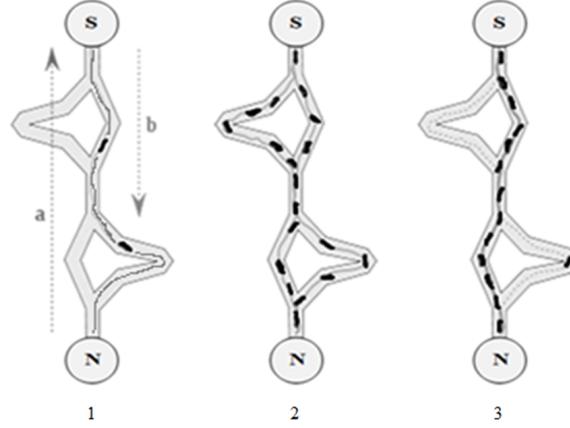


Figura 2: Figura ilustrativa del funcionamiento del algoritmo de las hormigas.

solución al problema.

Podemos representar el TSP a través de un grafo en el que las ciudades son los nodos (r, s) , los cuales estarán unidos por caminos pesados según el coste de realizar dicho camino $\delta(r, s)$. Además, cada nodo tendrá asociado un valor que denominamos feromona, $\tau(r, s)$, el cual da cuenta de lo atractivo que es dicho nodo para ser visitado, de este modo podemos evitar visitar nodos de forma repetida. Partimos por tanto de un grafo con las características descritas y un conjunto de hormigas las cuales resuelven el problema de forma individual partiendo desde nodos distintos. Desde cada nodo las hormigas se moverán al siguiente con probabilidad :

$$p_k(r, s) = \begin{cases} \frac{\tau(r, s) \cdot \mu(r, s)^\beta}{\sum_{u \in J_k(r)} \tau(r, s) \cdot \mu(r, s)^\beta} & \text{sí } s \in J_k(r) \\ 0 & \text{otro caso} \end{cases} \quad (5.1)$$

dónde μ es la inversa de la distancia $\delta(r, s)$, $J_k(r)$ es el conjunto de ciudades sin visitar por la hormiga k en la ciudad r y β es un parámetro configurable que determina la importancia que la hormiga da a el nivel de feromona contra la distancia que tiene que recorrer. Con el propósito de acumular una mayor cantidad de feromonas sobre los caminos más cortos la cantidad de feromonas en cada nodo se actualiza en varias ocasiones. Primeramente cada vez que una hormiga llega a un nodo cambia el nivel de feromona de este según:

$$\tau(r, s) \leftarrow (1 - \rho) \cdot \tau(r, s) + \rho \cdot \Delta\tau(r, s) \quad (5.2)$$

conocida como regla de actualización local, siendo $0 < \rho < 1$ un parámetro ajustable y $\Delta\tau(r, s)$ una función a optimizar según el problema. Siguiendo las expresiones 5.2 y 5.1 cada hormiga recorre todos los nodos del grafo. Esta actualización local hace, por lo general, que el nivel de feromona de cada nodo se reduzca haciéndolo menos atractivo para el resto de hormigas, se consigue así que estas exploren otros nodos. De esta forma nos aseguramos de que las hormigas nunca van a converger a un mismo recorrido de forma inmediata, lo cual es interesante pues cuantos más caminos exploremos mayor será la probabilidad de encontrar el óptimo. Una vez todas las hormigas han terminado el recorrido se vuelve a

actualizar, de forma global esta vez, el nivel de feromona de los nodos. Esta actualización global se hace siguiendo la expresión:

$$\tau(r, s) \leftarrow (1 - \alpha) \cdot \tau(r, s) + \alpha \cdot \Delta\tau(r, s) \quad (5.3)$$

con $0 < \alpha < 1$ el parámetro de decaimiento de las feromonas que da cuenta de la velocidad a la que estas se evaporan siguiendo la analogía con el caso de las hormigas reales. Por su lado;

$$\Delta\tau(r, s) = \begin{cases} \frac{1}{L_{gb}} & \text{sí } (r, s) \in \text{mejor ruta general} \\ 0 & \text{otro caso} \end{cases}$$

Siendo L_{gb} la longitud del camino total más corto encontrado hasta el momento. Con esta segunda actualización de las feromonas se completa una iteración del algoritmo, determinando así el nivel de feromona de cada nodo para la iteración siguiente. El valor concreto de los parámetros α , ρ , la función $\Delta\tau(r, s)$ en la actualización local o el número de hormigas se deberán ajustar según las características del problema que estemos tratando.

Si bien existen diversas implementaciones de este algoritmo, para este estudio se utilizó la librería [32], la cual se basa en lo apenas explicado para resolver un grafo dado. En concreto esta implementación del algoritmo sitúa a cada hormiga en un nodo aleatorio, siendo el número de hormigas un parámetro a elegir. Para elegir que nodo visitará cada hormiga en la iteración siguiente se evalúa el peso de cada nodo, w , a partir de su peso inicial, w_0 , su nivel de feromona, f , y los parámetros α y β también a elegir y que determinan la importancia del nivel de feromona y del peso del nodo. De este modo el algoritmo calcula $w = f^\alpha (\frac{1}{w_0})^\beta$. Guardando el peso de todos los nodos que aún no se han visitado crea otra lista con la acumulación de dichos pesos. Esto es, supongamos que tenemos una lista de pesos [1,2,3], la lista de pesos acumulados será [1,3,6]. A partir de estas listas el algoritmo calcula la intersección entre la lista de acumulados y la suma de estos multiplicada por un número aleatorio entre 0 y 1 quedándose con la posición de la intersección. Es decir, en nuestro caso anterior la suma sería 10, si al multiplicarla por el número aleatorio obtenemos el 5 este número interseca con la lista entre el 3 y el 6 de modo que nos quedaríamos con la tercera posición (el índice y no el valor). De la lista de pesos de los nodos por visitar se toma la posición apenas obtenida y el nodo cuyo peso se corresponda a este valor será el siguiente que la hormiga visite. De esta forma tienen más probabilidad de ser elegidos los nodos con mayor peso pero al incluir la componente aleatoria también es posible visitar los demás.

6 Algoritmos de machine learning

Los físicos, están comúnmente familiarizados con el tratamiento y procesado de grandes cantidades de datos. Los experimentos llevados a cabo en el LHC o en proyectos astronómicos como el Sloan Digital Sky Survey (SDSS) generan constantemente petabytes y terabytes de datos de estudio. Si bien es cierto que debido

a la precisión que suelen tener los experimentos físicos, puede resultar chocante el tratar con grandes conjuntos de datos dispares o con excesivo ruido [33]. En otras ramas del conocimiento, la cantidad de datos con los que tratar ha crecido también hasta el punto de requerir un procesado automatizado. Es en este contexto donde los algoritmos de machine learning nos permiten, a través de la detección de patrones en dichos datos, tratarlos, clasificarlos o simular datos nuevos entre otras posibilidades. En el caso concreto de la física los investigadores han estado investigando y desarrollando las aplicaciones de esta rama de la inteligencia artificial en sus proyectos, en astrofísica las aplicaciones son extensas [34], [35]. En el campo de la física estadística se usan algoritmos de inteligencia artificial para detectar transiciones de fase en el modelo de Ising bidimensional [36], en la física cuántica también encontramos variedad de aplicaciones como la caracterización de estados cuánticos [37] o correlación de error en información cuántica [38]. Estos son solo algunos ejemplos de los muchos que muestran como los físicos están aprovechando las ventajas de los algoritmos de machine learning en sus investigaciones, [39]. Además, los físicos tienen un perfil adecuado para entender y adoptar estos algoritmos puesto que algunas de las técnicas y conceptos usados en machine learning tienen su origen en problemas físicos; métodos Monte-Carlo o métodos variacionales e incluso algunos modelos se construyen a partir de conceptos como la optimización energética, o el estudio de la eficiencia termodinámica de los métodos de aprendizaje [40].

Antes de poder estudiar en detalle los datos obtenidos en cualquier experimento, es conveniente clasificar de donde provienen dichos datos. Así en un acelerador de partículas conviene saber si los datos que estamos tratando provienen de un electrón, un protón, o de una u otra interacción... mientras que en una observación astronómica conviene saber si estamos observando un fenómeno u otro o diferenciar los datos de interés de posible ruido para poder determinar que modelo aplicar sobre los datos. Esta primera tarea de clasificación es por tanto crucial cuando el origen de los datos es incierto. Se presentan a continuación algunos de los algoritmos de machine learning comúnmente utilizados para este fin.

6.1 K-Nearest-Neighbor, KNN

Pese a su simplicidad, el algoritmo de los K vecinos más cercanos, es de los más usados en el ámbito del machine learning pues demuestra gran eficacia en problemas tanto supervisados como no supervisados. La idea detrás de este algoritmo es la de clasificar los datos según lo parecidos (ceranos) que sean a otros datos del conjunto de entrenamiento (vecinos). Para tal fin, necesitamos guardar todo el conjunto de entrenamiento y recorrerlo durante la evaluación para encontrar los conjuntos de datos similares. Formalmente, supongamos que partimos de un dominio χ y una métrica que devuelve la distancia entre puntos del dominio $\rho : \chi \times \chi \rightarrow \mathbb{R}$. Para el caso común de $\chi = \mathbb{R}^d$ se suele tomar como métrica la

distancia euclídea;

$$\rho(x, x') = \|x - x'\| = \sqrt{\sum_{i=1}^d (x_i - x'_i)^2}$$

En tal caso, para un conjunto de datos de entrenamiento etiquetados $S = (x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$ definimos $\pi_1(x), \pi_2(x), \dots, \pi_m(x)$ como el reordenamiento del conjunto según la distancia $\rho(x, x_i)$ para $i < m$ y $x \in \chi$, es decir, $\rho(x, x_{\pi_i(x)}) \leq \rho(x, x_{\pi_{i+1}(x)})$. Para cada punto $x \in \chi$ el algoritmo devuelve la etiqueta mayoritaria del conjunto $\{y_{\pi_i(x)} : i \leq k\}$.

6.2 Support Vector Machine, SVM, y métodos kernel

Otra de las herramientas de uso más extendido en la aplicación de algoritmos de machine learning es support vector machine (SVM). Su principal función es la predicción lineal incluso en conjuntos de datos multidimensionales. De forma esquemática este algoritmo busca separar linealmente el conjunto de datos mediante un hiperplano. Para llevar a cabo esta tarea de forma óptima el algoritmo busca el hiperplano que, aun separando los datos, quede más alejado de ellos. Para los algoritmos que si sean linealmente separables, Hard-SVM, el conjunto de datos quedará dividido perfectamente, en cambio, si los datos no son linealmente separables, Soft-SVM, el algoritmo tratará de optimizar la cantidad de datos que quedan divididos en otro subespacio. Consideremos el conjunto de datos $S = (x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$ tales que $x_i \in \mathbb{R}^d$ y por simplicidad definimos las etiquetas como $y = \{-1, 1\}$. Este conjunto será linealmente separable si existe un semiplano (\vec{u}, b) tal que $y_i = \text{sign}(\langle \vec{u}, x_i \rangle + b) \quad \forall i = 1, \dots, m$, es decir, todos los datos con la misma etiqueta quedan a un mismo lado del semiplano. El hiperplano que define esta separación vendrá definido como $L = \{\langle \vec{v}, \vec{u} \rangle + b = 0\}$ Definimos como margen a la mínima distancia entre L y el conjunto de datos $d(\vec{x}, L) = \min\{\|\vec{x} - \vec{v}\| : \vec{v} \in L\}$ que para $\|u\| = 1$ queda $d(\vec{x}, L) = |\langle \vec{u}, \vec{x} \rangle + b|$. Los elementos del conjunto de datos más cercanos al hiperplano se renombran como vectores soporte. En general, dado el conjunto S , el algoritmo resuelve:

$$(\vec{u}_0, b_0) = \underset{(\vec{u}, b)}{\text{argmin}} \|\vec{u}\|^2 \quad ; \quad y_i(\langle \vec{u}, \vec{x} \rangle + b) \geq 1 \quad \forall i \quad (6.1)$$

Finalmente, el algoritmo devuelve $\hat{u} = \frac{\vec{u}_0}{\|\vec{u}_0\|}$, $\hat{b} = \frac{b_0}{\|\vec{u}_0\|}$. Los vectores que definen el margen, vectores soporte, son los únicos que determinan \vec{u}_0 , es decir, existe un conjunto de coeficientes a_1, a_2, \dots, a_m tales que $\vec{u}_0 = \sum_{i \in I} a_i \vec{x}_i$ donde I es el conjunto de índices de los vectores soporte.

En el caso de que el conjunto de datos no sea linealmente separable, Soft-SVM, relajamos la condición expuesta en la expresión 6.1 y cuantificamos cuanto la estamos violando. Para ello introducimos el conjunto de variables auxiliares $\xi_1, \xi_2, \dots, \xi_m$ siendo $\xi_i \geq 0$ con las que minimizar la violación de la restricción. De esta forma ahora, a partir del conjunto de datos S y un parámetro λ para regular el balance entre la solución óptima y la violación de la restricción, devolviendo

nuevamente \hat{u}, \hat{b} , el algoritmo resolverá:

$$(\vec{u}_0, b_0) = \min_{(\vec{u}, b, \xi)} \left(\lambda \|\vec{u}\|^2 + \frac{1}{m} \sum_{i=1}^m \xi_i \right) \quad ; \quad y_i(\langle \vec{u}, \vec{x} \rangle + b) \geq 1 - \xi_i \quad \forall i \quad (6.2)$$

No obstante, hay conjuntos de datos para los cuales no es posible aplicar directamente esta estrategia. Pensemos, por ejemplo, en un conjunto de puntos χ sobre una línea tal que los datos en el centro tienen una etiqueta distinta a los demás. A priori, no es posible separar linealmente dicho conjunto. Una forma de solucionar este problema podría ser definir una función $\Psi : \chi \rightarrow \mathcal{H}$ que mapease los datos a un espacio de Hilbert en el que fuesen linealmente separables, aplicar el algoritmo sobre dicho espacio y mapear el resultado de nuevo al espacio original. Así, en el ejemplo anterior podemos mapear cada dato según $\Psi(x) = x^2$, ahora los datos quedan distribuidos en una parábola sobre la que podemos separarlos de forma lineal. Este método, sin embargo, puede ser computacionalmente muy complejo ya que el nuevo espacio por lo general será de dimensionalidad mayor que el original de modo que se suelen considerar otras alternativas como los métodos kernel.

Por kernel, en este caso, entendemos el producto interior en el espacio de atributos definiendo la función kernel $K(x, x') = \langle \Psi(x), \Psi(x') \rangle$. La principal ventaja de este método es que es posible aplicar el algoritmo sobre los valores de la función kernel sin necesidad de representar explícitamente los datos en el espacio de atributos. La minimización que tendrá que hacer el algoritmo ahora será $\min_{\vec{u}} (f(\langle \vec{u}, \Psi(x_1) \rangle, \dots, \langle \vec{u}, \Psi(x_m) \rangle) + R(\|\vec{u}\|))$ donde $f : \mathbb{R}^m \rightarrow \mathbb{R}$ es una función genérica y $R : \mathbb{R}_+ \leftarrow \mathbb{R}$ una función monótona no decreciente. Puesto que existe un vector $\vec{a} \in \mathbb{R}^m$ tal que $\vec{u} = \sum_{i=1}^m a_i \Psi(x_i)$ a partir de la definición de la función kernel la función a minimizar quedará:

$$\min_a \left(f\left(\sum_j a_j K(x_j, x_1), \dots, \sum_j a_j K(x_j, x_m)\right) \right) + R \left(\sqrt{\sum_{i,j} a_i a_j K(x_i, x_j)} \right) \quad (6.3)$$

Pese a que este método resuelve el problema de los conjuntos de datos no linealmente separables debemos decidir que función kernel usar. Esta puede variar mucho según el problema a resolver. No obstante una de las opciones más interesantes y la usada en este trabajo es el kernel Gaussiano, la función base radial, RBF por sus siglas en inglés. Esta, viene definida como

$$K(x, x') = e^{-\frac{\|x-x'\|^2}{2\sigma^2}} \quad (6.4)$$

Aquí, la desviación estándar σ regula la anchura del margen. Este kernel tiene la ventaja de que con los parámetros adecuados debería ser capaz de delimitar todos datos correctamente, aunque esto no suele interesar en los problemas de machine learning pues estaríamos sobre-entrenando el modelo.

6.3 Neural Networks, NN

Inspiradas en las neuronas las redes neuronales son modelos no lineales usadas para el aprendizaje tanto supervisado como no supervisado. La unidad de computación básica de una red neuronal son las neuronas, esto dota a la red de una estructura de grafo en la que cada neurona representa un nodo y está unida con otras neuronas mediante vértices. Cada neurona recibe la suma ponderada de los resultados calculados por las neuronas a las que está conectada. Esto es, cada neurona i toma un vector $\vec{x} = (x_1, \dots, x_d)$ y devuelve un escalar $a_i(\vec{x})$. Normalmente en las redes neuronales encontramos varias neuronas organizadas por capas, así, la primera capa se considera la de entrada, la última la de salida, y el resto las capas ocultas. Pese a que a_i variará según el propósito de la red, su función siempre será la de pesar los datos o los resultados que obtenemos de cada capa. Para esto se aplica sobre la entrada primeramente una función lineal $z(\vec{x})$ y sobre el resultado de esta una función no lineal $\sigma(z)$. La función $z(\vec{x})$ que suelen tomar la mayoría de redes es $z_i = \vec{w}_i \cdot \vec{x} + b_i = \vec{x}^T \cdot \vec{w}_i$ donde $\vec{w}_i = (w_1, \dots, w_d)$ es el peso de cada neurona sobre la que aplicamos una cierta restricción b_i . Por su parte la función no lineal dependerá en gran medida del tipo de problema que estemos resolviendo, aunque su desarrollo detallado excede el propósito de este trabajo algunos ejemplos recurrentes son la función escalón, la función Tanh o la función sigmoideal. Son sin embargo de especial interés por su posterior uso en este estudio, las funciones 'RELU' y 'sigmoid'. La función sigmoid tiene la forma:

$$\sigma(z)_i = \frac{1}{1 + e^{-z_i}}$$

Esta función se usa recurrentemente en la capa de salida en problemas de clasificación múltiple siendo K el número de clases a clasificar. Por su parte la función RELU tiene la forma $\sigma(z)_i = \max(0, z_i)$.

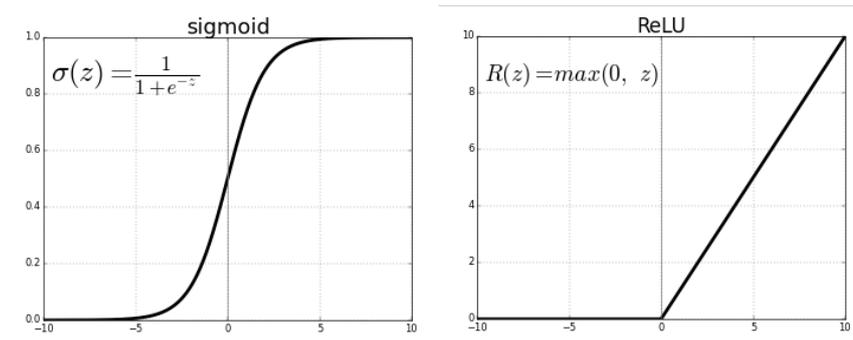


Figura 3: Funciones de activación sigmoid y Relu.

Para entrenar los algoritmos supervisados necesitamos además una función perdida que optimizar durante el proceso de entrenamiento. Esto es, una vez obtenidos los datos de la capa de salida comparamos y calculamos cuanto se alejan de su valor real y reajustamos los pesos de la red para volver a pasar los datos por ella en una nueva iteración. Esta función de coste se suele optimizar mediante un algoritmo de descenso de gradiente o alguna de sus variantes como 'adam'

[41], utilizada en este trabajo. En concreto, para problemas de clasificación se suele tomar como función de coste la función 'crossentropy' junto con la función de activación sigmoid.

$$CE = -\log \left(\frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \right)$$

7 Metodología

Debido a la gran cantidad de datos que se prevé que el observatorio CTAO registre, se proponen en este trabajo algunas alternativas tanto para la clasificación de los datos como para el guardado de estos. Dada tal cantidad de datos a tratar, se optó por estudiar una primera aproximación de la clasificación de estos mediante diversos algoritmos de machine learning. Por la naturaleza clasificatoria del problema se aplicaron los algoritmos de k-vecinos más cercanos, vectores soporte y redes neuronales. Dichos algoritmos se implementaron con sus parámetros por defecto pues excede los propósitos de este trabajo hacer un estudio intensivo sobre bajo qué parámetros o métricas dan mejores resultados. Tras la comparativa entre algoritmos se estudió un modo alternativo de guardar los datos para su posterior procesado. Alguna colaboración como DUNE (Deep Underground Neutrino Experiment) ha propuesto su algoritmo, PANDORA, para la reconstrucción gráfica de los eventos que observan en sus instalaciones [42]. No obstante, en este trabajo se propone una aproximación diferente y novedosa, expresar los datos registrados por los telescopios de CTAO a través de grafos. Se proponen por tanto, tras un análisis de los datos, tres maneras distintas de construir dichos grafos de forma heurística mediante el algoritmo de las hormigas. Se busca de esta forma encontrar la representación óptima de estos. A partir de la configuración de datos encontrada por las hormigas se vuelven a probar los modelos de clasificación sobre este nuevo conjunto de datos y se comprueba si dicha representación supone una reducción del espacio en memoria.

8 Exploración de los datos

Los datos de estudio de este trabajo pertenecen a un conjunto de observaciones realizadas por el primer prototipo de LST (Large Size Telescope) del observatorio CTAO situado en el observatorio de El Roque de los Muchachos capaz de detectar eventos a partir de los 20 GeV con sus 1855 píxeles. El conjunto de datos facilitado estaba compuesto, entre especificaciones del propio telescopio, de los registros correspondientes a 41083 eventos. Los principales datos a tener en cuenta de fueron los valores de carga de fotoelectrones registrados por cada píxel en cada evento puesto que, como hemos visto, nos permiten, junto con el tiempo de registro, determinar la naturaleza del mismo; Cada evento contaba ya con una clasificación según su origen, en total las etiquetas eran 'cosmic' para eventos pertenecientes a rayos cósmicos, 'flatfield' para registros correspondientes a la iluminación artificial del telescopio con fines de calibración y 'pedestal' para registros donde solo

se midió ruido. El conjunto de datos también proporcionaba los datos relativos a la limpieza de estos especificando qué píxeles contenían datos de ruido además de otros parámetros relativos a la discriminación del tipo de rayo cósmico y el cálculo de la energía y ángulo de incidencia de este. Para este estudio, no obstante, solo se tuvieron en cuenta los datos relativos a la carga de fotoelectrones (carga de phe) de cada píxel y el tipo de evento en el que estaba clasificado.

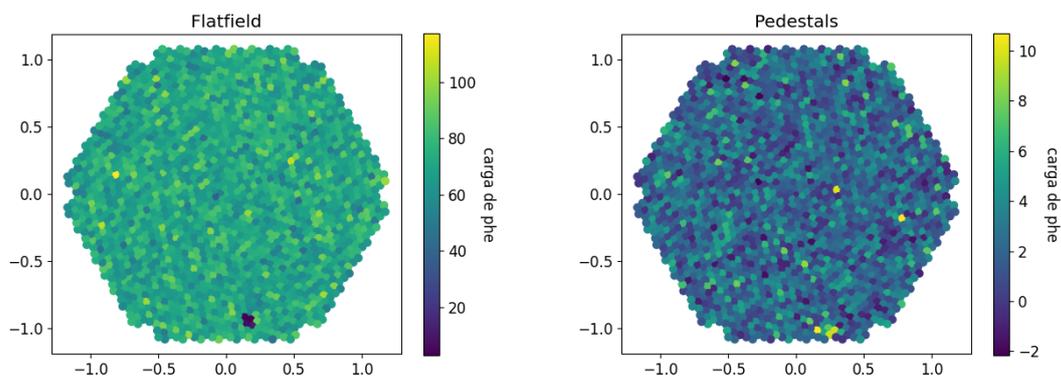


Figura 4: Visualización de eventos de flatfield y pedestal sobre el telescopio

Los eventos registrados como 'flatfield' presentan una carga de entre 80 y 100 phe por píxel. Estos eventos se generan de forma artificial mediante focos de luz controlada para medir el correcto funcionamiento de los píxeles. De hecho, podemos apreciar en la figura 4 como para el evento de flatfield hay algunos píxeles que no se iluminaron, indicando su mal funcionamiento. Por su parte, los eventos etiquetados como 'pedestals' presentan una carga próxima o incluso inferior a cero, siendo su carga media máxima entorno a los 12 phe por píxel. En estos eventos no se registra ningún evento cósmico, tan solo el ruido que llega al telescopio. Al igual que en los eventos 'flatfield' en los 'pedestals' no se distingue, como muestra la figura 4, patrón alguno sobre los píxeles.

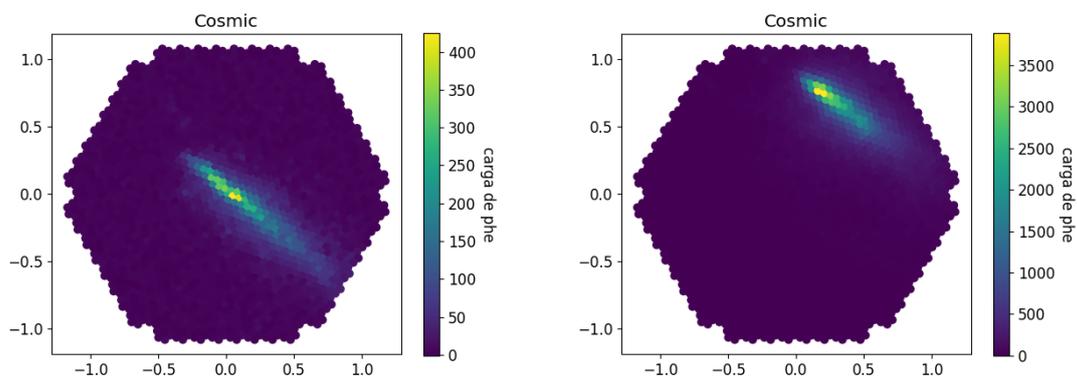


Figura 5: Visualización de eventos producidos clasificados como 'cosmic' sobre el telescopio

El rango de valores en la carga de phe por píxel de los eventos cósmicos, registrados como 'cosmic' es mayor pues dichos eventos corresponden rayos gamma de alta energía o rayos cósmicos que generan fotones mucho más energéticos. Así mismo, a diferencia del resto de eventos ya mencionados, estos no se registran de manera general en todos los píxeles del telescopio, si no que se localizan en un número determinado de píxeles. En la figura 5 se aprecian eventos que pudieran ser causados por rayos gamma debido a la distintiva proyección elipsoidal del cono de emisión de radiación cherenkov. No obstante la determinación exacta del tipo de rayo originario no se puede determinar a simple vista de las gráficas. Se observa además como, ambos eventos, muestran valores máximos de carga de phe por píxel dispares. Esto, podría sugerir que la causa originaria de cada rayo fuera distinta.

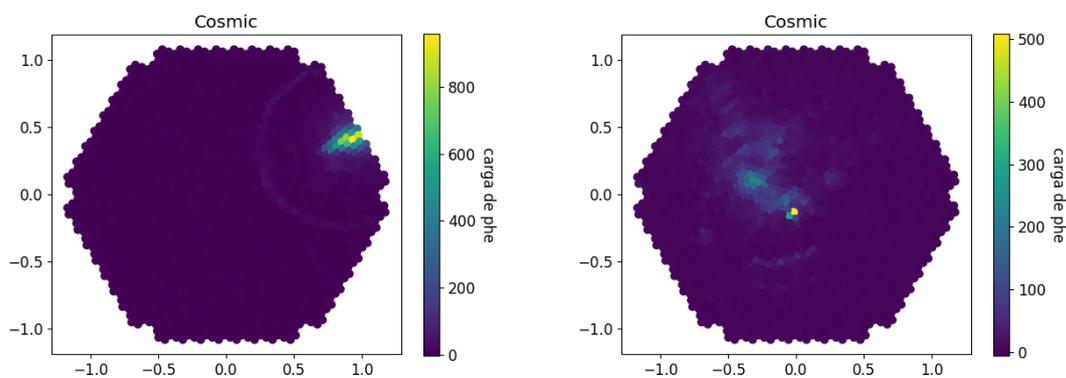


Figura 6: Visualización de eventos clasificados como 'cosmic' sobre el telescopio

Puesto que los rayos cósmicos son partículas cargadas, la cascada que crean al entrar en la atmósfera será distinta en función de la partícula concreta que la ocasiona. Es por esto que la visualización de estos eventos sobre el telescopio tiene una mayor variedad de patrones. No obstante, dentro de la miríada de posibilidades hay algunos patrones como el que deja el pión que sí tienen patrones reconocibles, en este caso por su característico anillo entorno al punto de medida central. Otros sin patrón aparente se muestran como una nube de píxeles iluminados de forma no definida, causa de la variedad de productos cargados producidos en la cascada.

En total, el conjunto de datos cuenta con 781 eventos registrados como 'flatfield' (calibración artificial), 774 como 'pedestals' observaciones de ruido y 39528 como 'cosmic' pertenecientes a eventos cósmicos. No obstante, aunque los eventos cósmicos sean mayoritarios, estos contienen también píxeles con valores de carga de phe asociados a la medición de ruido. Cuando se registra un evento cósmico no todos los píxeles del telescopio registran fotones de radiación Cherenkov como se puede apreciar en las figuras 6, 5. La representación de todos los datos de carga de phe en función del píxel que los registró pone de manifiesto este hecho. Los eventos de 'flatfield' se activan aproximadamente en un ratio de cien por segundo, puesto que el conjunto de datos tiene 781 eventos registrados de este tipo, el conjunto de datos corresponde a algo menos de 8 s de medición. Esto explica

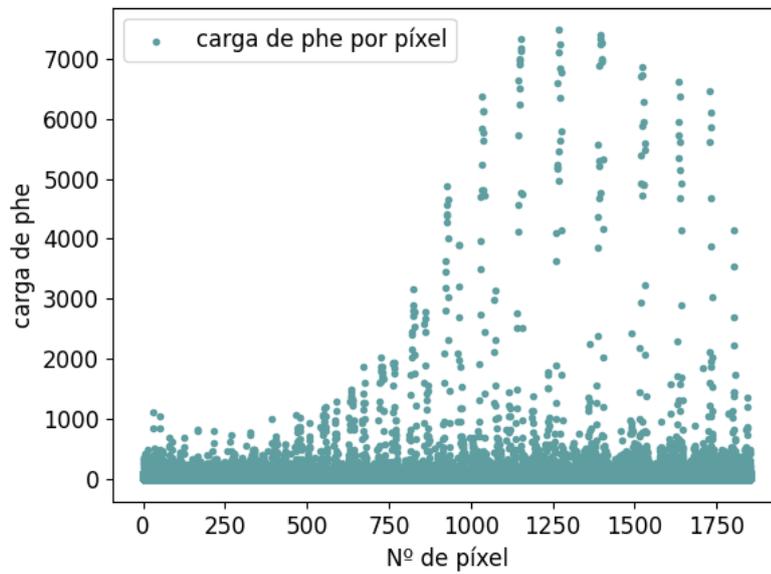


Figura 7: Carga de phe medida por cada píxel a lo largo del todo el conjunto de datos

la aparente diferencia de probabilidad entre píxeles de observar eventos de altas energías que se podría deducir de la figura 7. Sin embargo, todos los píxeles tienen la misma probabilidad de captar eventos de alta energía, es decir, bajo una cantidad suficiente de datos la figura 7 sería una distribución plana. La mayoría de los valores son cercanos al cero, si bien algunos correspondientes a eventos más energéticos llegan a superar una carga de phe de 7000.

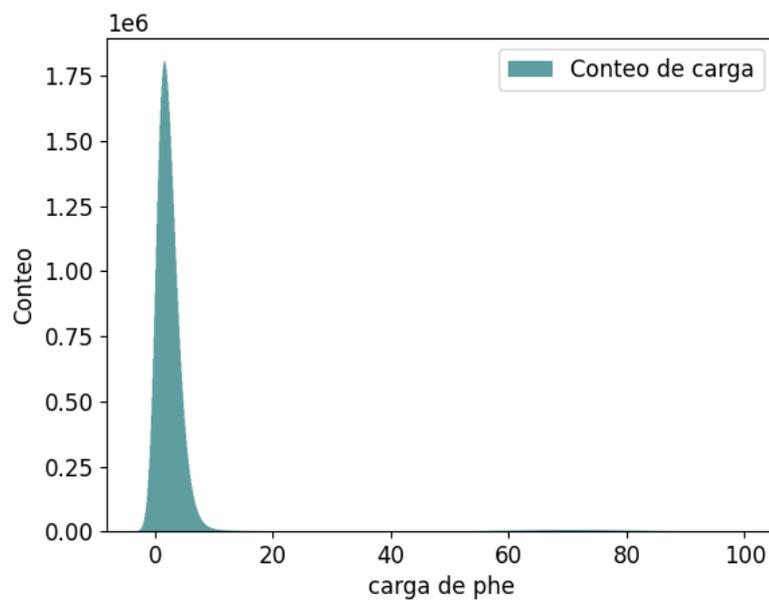


Figura 8: Conjunto de datos con carga de píxel inferior a 100 phe.

Esta amplitud en el rango de valores de nuestros datos y, sobre todo, la disparidad en su densidad, hace que una representación del número de cuentas de cada valor de carga de phe sea visualmente ineficiente. Cada evento consta de 1855 datos, teniendo en cuenta los 41083 eventos de los que disponemos el total de datos individuales suma a más de $76 \cdot 10^6$. De todos estos datos solo algo más de 10000 son correspondientes a valores de carga de phe por encima de 100. Es por esto que para poder poner en perspectiva el conjunto de datos se consideraron solo aquellos valores por debajo de este límite. Incluso mostrando los valores de carga menores a 100 pudiera parecer, observando la figura 8, que este límite podría disminuirse aún más pues la mayoría de los datos se concentran en valores cercanos al cero. Haciendo uso de los datos relativos a la limpieza de las imágenes se discriminó de entre estos datos, los correspondientes a ruido de los que realmente tienen mayor interés para su estudio.

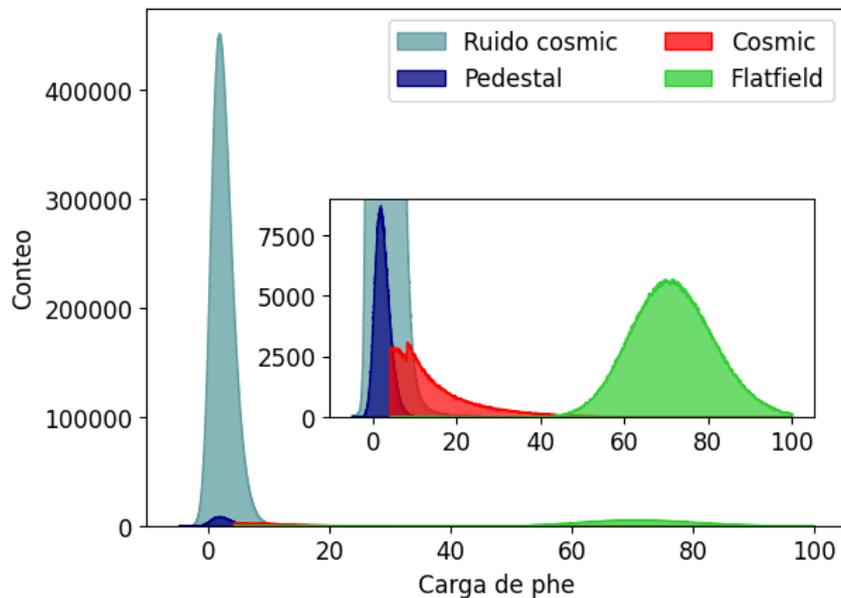


Figura 9: Conjunto de datos de ruido junto a los datos útiles para valores de carga de píxel inferiores a 100 phe.

Observando la figura 9 se pone de relieve la diferencia entre la cantidad de datos pertenecientes a ruido y los que realmente son de utilidad. Al tener en cuenta que estos datos provienen de un solo telescopio y que en el futuro CTAO estará formado por 64 de ellos se aprecia más claramente la necesidad de discriminar de manera eficiente los datos útiles de los de ruido. Para una mayor comodidad en el posterior tratamiento de datos se consideraron los valores de ruido con una carga inferior a 11 phe por píxel y se observó que el ruido se distribuía asemejándose a una función gaussiana del tipo:

$$g(x) = A_0 e^{-\frac{(x-\mu)^2}{2\sigma^2}} + B_0$$

Siendo μ y σ la media y la desviación del ruido respectivamente y A_0 y B_0 constantes. Se ajustó el conjunto de datos de ruido mediante esta función obteniéndose los parámetros mostrados en la tabla 1.

Parámetro	Ajuste
μ	$1.642 \pm (8.0 \cdot 10^{-3})$
σ	$1.982 \pm (7.0 \cdot 10^{-3})$
$A_0 \cdot 10^3$	258.57 ± 1.03
$B_0 \cdot 10^3$	3.28 ± 0.53

Tabla 1: Parámetros de ajuste de los datos de ruido por una función gaussiana.

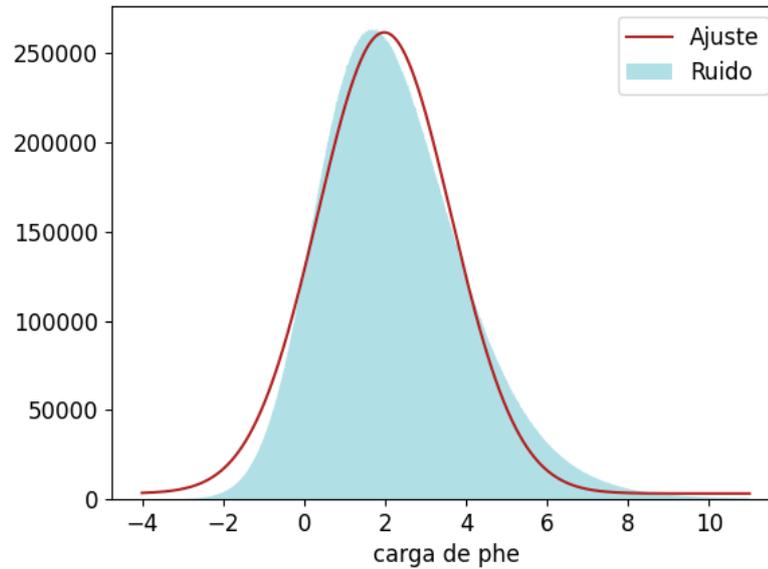


Figura 10: Ajuste de los datos de ruido a una función gaussiana.

Para determinar con mayor precisión la bondad del ajuste, se realizó el test de χ^2 para el cual se obtuvo $\chi_v^2 = 0.07$. Valor que indica que los datos se ajustan bien a una gaussiana. χ_v^2 se obtuvo como $\chi_v^2 = \frac{\chi_{exp}^2}{n-p}$ siendo n la extensión del número de datos, p los grados de libertad asociados, 4 en este caso por tener 4 parámetros de ajuste y χ_{exp}^2 , siendo ϵ_i la desviación típica de los valores, tal que $\chi_{exp}^2 = \sum_{i=1}^n \frac{(y_{i,obs} - y_{i,ajuste})^2}{\epsilon_i}$. Es claro de la figura 10 que la gaussiana no está centrada en el cero, este hecho encuentra su explicación en el método que desde CTAO utilizan para procesar los datos de ruido. Por cada píxel a lo largo de la medida de un evento filtran el ruido registrado solamente los valores máximos de carga de phe, es decir, los datos de ruido están sesgados. De no ser así, la distribución de ruido estaría centrada en el cero.

9 Representación mediante grafos

Se propone en este trabajo la representación mediante grafos, usando la librería networkx [43], de los datos obtenidos por los IACTs de CTAO como método a considerar para la necesaria reducción del almacenamiento de estos. Debido al

alto volumen de datos y su consecuente coste de procesamiento que espera tener CTAO, la representación de los mismos mediante grafos se presenta como opción a tener en cuenta para aliviar carga en almacenamiento y el procesamiento de datos. Las posibilidades para crear grafos a partir de los datos obtenidos por un IACT son numerosas, algunas opciones a tener en cuenta podrían ser; utilizar los datos de carga de phe por píxel como nodos e interconectarlos asignando a cada nodo el mismo peso o asignando el tiempo de registro como peso de cada nodo, utilizar los datos temporales como nodos con peso la carga de phe de cada píxel, representar en lo anterior el ruido mediante su media μ y su desviación σ obtenidas en el ajuste de la figura 10, representar el ruido mediante otro grafo que conforme uno de los nodos, representar el tiempo mediante un grafo que conforme uno de los nodos... Cada opción facilita tratar la información contenida en el grafo de distintas formas.

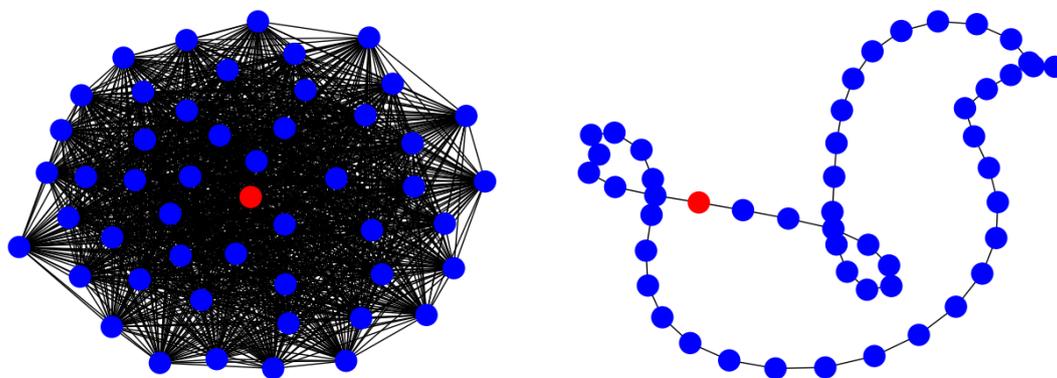


Figura 11: Grafo antes y después de aplicar el algoritmo de las hormigas. Cada nodo representa un valor de carga de phe por píxel mientras que el nodo de color rojo esta formado por el ruido expresado como el conjunto (μ, σ) .

En este estudio, debido a las limitaciones del mismo, se consideran tres posibilidades. En la primera, los nodos del grafo están formados por los datos marcados como de interés mientras que el ruido se representa mediante μ y σ que forman cada una un nodo respectivamente, todos los nodos están interconectados y su peso es inversamente proporcional a la distancia entre ellos. La segunda, se diferencia de la primera en que los parámetros de ajuste del ruido se representan ambos en un único nodo. En la tercera, de forma similar a la segunda, se aúnan todos los datos de ruido en un grafo independiente que conforma un nodo y se interconecta con el resto de datos al igual que en el primer método. Los grafos así expresados son aún poco eficientes para su tratamiento pues al estar todos los nodos interconectados forman un conjunto de datos todavía mayor que el de partida. Es por esto que recurrimos al algoritmo de las hormigas para su procesamiento. Las 'hormigas' encuentran de forma heurística la forma óptima de interconectar los nodos de forma que cada nodo solo esté conectado con otros dos.

Las figuras, 11 y 12 representan el mismo evento. Aunque visualmente las representaciones sean parecidas cada figura se corresponde con una interpretación distinta de la construcción del grafo. Es más, puesto que el algoritmo de las hor-

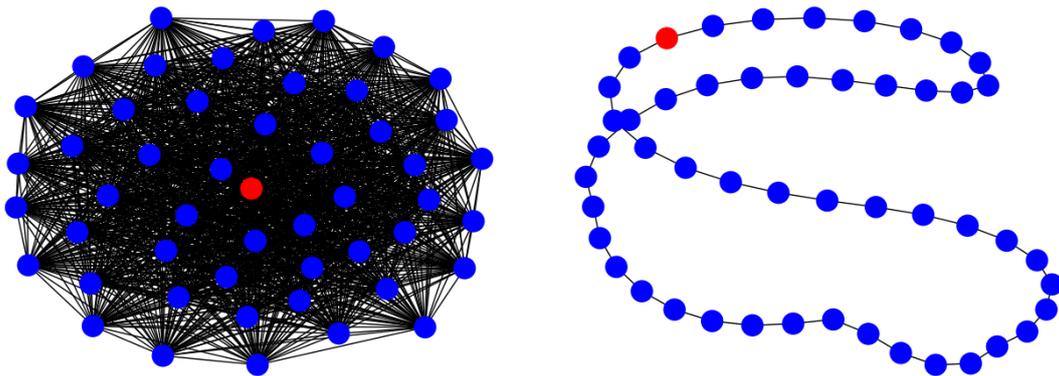


Figura 12: Grafo antes y después de aplicar el algoritmo de las hormigas. Cada nodo representa un valor de carga de phe por píxel mientras que el nodo de color rojo esta formado es otro grafo formado por todos los datos de ruido del evento.

migas es un algoritmo heurístico, si volviésemos a correr este sobre los grafos originales no tendríamos por que obtener las mismas simplificaciones. Este hecho depende también de la propia implementación del algoritmo, pues, por defecto, el nodo desde el que las hormigas comienzan su recorrido es aleatorio. Para la visualización de la primera configuración se escogió un evento distinto por su posterior implementación.

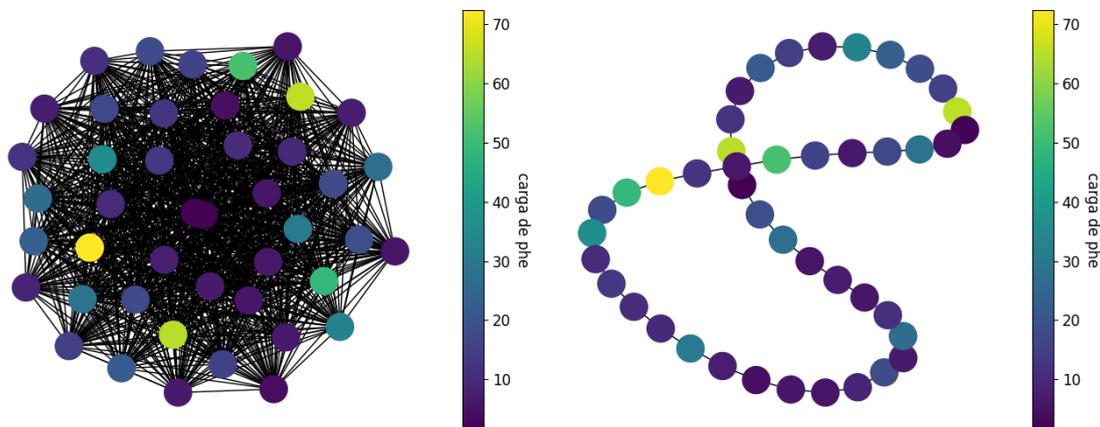


Figura 13: Grafo antes y después de aplicar el algoritmo de las hormigas. Cada nodo representa un valor de carga de phe por píxel.

Debido a la elección aleatoria del nodo inicial, el algoritmo de las hormigas generará, a priori, un grafo distinto como apreciamos en las figuras 14 y 13. El algoritmo de las hormigas se aplicó sobre los grafos originales en su versión más básica. Se usaron cinco hormigas que iteraron una vez sobre todos los nodos, así mismo, el nivel de feromonas y su desvanecimiento en cada conexión se mantuvo por defecto tal y como venía implementado en la librería usada. Un estudio más exhaustivo sobre como habría afectado la variación de cada uno de los parámetros modificables del algoritmo a la resolución del grafo excedía los propósitos del presente estudio. No obstante con el fin de subsanar la aleatoriedad del no-

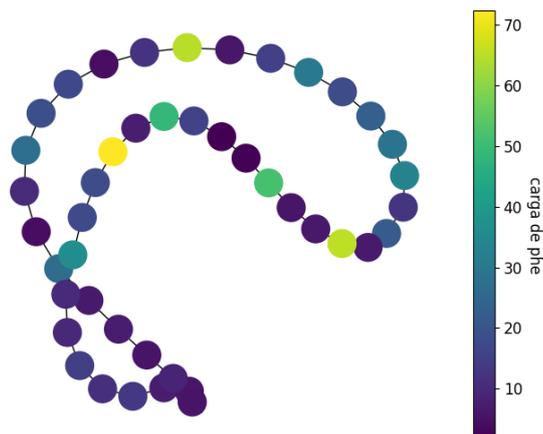


Figura 14: Segundo resultado devuelto por las hormigas para el evento mostrado en la figura 13

do inicial y poder comprobar si las hormigas convergían a un mismo camino se seleccionó un evento con pocos nodos y se hizo iterar a un número de hormigas igual al número de nodos durante el doble de iteraciones que de nodos.

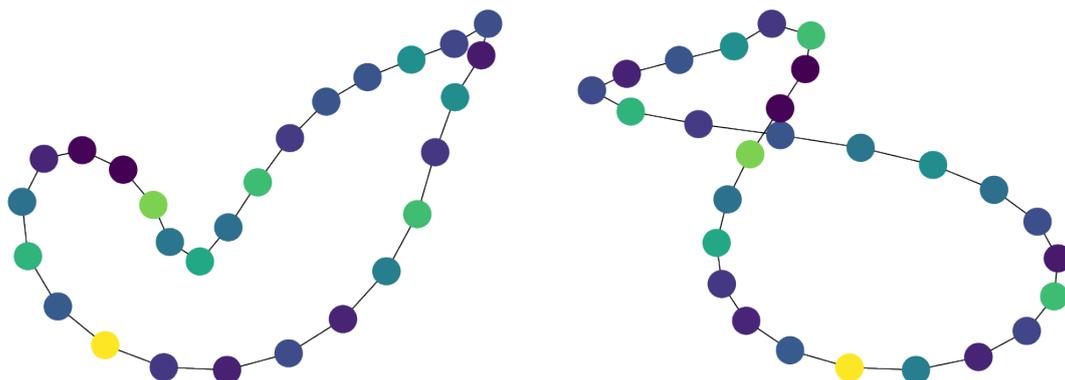


Figura 15: Grafos generados con 44 hormigas iterando 88 veces para compensar la aleatoriedad inicial.

Puesto que se usaron las mismas hormigas que nodos el punto inicial de cada una dejó de importar pues cada hormiga es indistinguible, además al aumentar el número de iteraciones a el doble del número de nodos nos aseguramos que las hormigas pudieran iterar suficiente para converger. Observando el nodo de mayor energía, amarillo, en la figura 15 vemos que las hormigas determinaron caminos distintos para el mismo evento en las mismas condiciones. No obstante, se pueden apreciar algunos patrones como que los dos nodos más oscuros, pertenecientes al ruido, aparecen juntos o que los nodos más energéticos no parecen agruparse si no que se intercalan con los de menor energía. Esto se asemeja a lo observado en la figura 7 donde los píxeles de mayor energía parecen espaciados entre sí mientras que, sobre los píxeles reales del telescopio, sabemos que los puntos de alta carga suelen aparecer juntos. La explicación de esta separación en el caso de la figura de carga por píxel se debe al orden de lectura de los datos de

cada píxel el cual es concéntrico partiendo del píxel central. Bajo suficientes iteraciones las hormigas podrían tender a una configuración similar de los datos.

9.1 Comparativa de recursos

Uno de los objetivos por los que se ha optado por una representación mediante grafos de los datos es la reducción de espacio en memoria que, a priori, esto conllevaría. Como se mencionó en la sección 4 uno de los métodos para determinar si el evento observado pertenece o no a un rayo gamma consiste en ajustar la representación de los datos obtenida, como la mostrada en las figuras 5, mediante la sección transversal del cono de emisión de la luz Cherenkov. La representación mediante grafos elegida no parece a simple vista permitir esta clasificación, no obstante, corroborar esta afirmación precisaría de un estudio más detallado al respecto. Además, en caso de no ser así otras representaciones alternativas en la construcción del grafo si podrían permitirla. Es por ello que se ha optado comparar el coste de almacenar los datos en grafos, tal cual mostrados en figuras como la 15, con respecto a almacenarlos mediante imágenes como las mostradas en las figuras 5, 6, 4. No obstante, los datos a partir de los que se han formado los grafos han sido preprocesados previamente para reducir la presencia de los valores de ruido. No sería por tanto indicativo compararlos con la representación de los datos sin procesar pues la simplificación del ruido a su media y desviación ya supone una reducción de datos. Para poder obtener una comparación significativa se representó la carga por píxel de algunos eventos con la salvedad de que a los píxeles clasificados como ruido se les asignó el valor 0, técnica también usada por el equipo de CTAO. Se comparó entonces el espacio en memoria de ambas representaciones para un conjunto de 11 eventos aleatorios que incluyera tanto eventos de 'pedestal' como 'cosmic'. Además dado el interés de los eventos 'cosmic' con altos valores de energía se realizó también la comparativa para 6 eventos 'cosmic' de estas características no incluidos en el conjunto anterior.

Píxel (KB)	Grafo (KB)	Píxel(cosmic)(KB)	Grafo(cosmic)(KB)
76.59	33.42	76.76	47.9

Tabla 2: Espacio medio en memoria de la representación mediante píxeles y mediante grafos de un subconjunto aleatorio de datos de CTAO y otro solo con datos de eventos 'cosmic'.

La representación mediante grafos supone una clara reducción del espacio medio en memoria de los datos respecto a la representación de los píxeles, si bien para los eventos de 'cosmic' la reducción es menor, en concreto se redujo un 56 % y un 38 % el espacio en memoria respectivamente. Los grafos de los eventos 'cosmic' son más pesados que el resto pues están, por lo general, compuestos por una cantidad mayor de nodos. No obstante, para la escala de la cantidad de datos que gestiona y prevé gestionar CTAO dicha reducción puede resultar muy significativa.

10 Rendimiento de los algoritmos de ML

Junto con la necesidad de un método de almacenamiento de los datos más eficaz, CTAO precisa también de algoritmos capaces de clasificar de forma automática tales cantidades de datos. Los algoritmos de machine learning juegan por tanto, un rol clave en esta tarea pues como ya se explicó en la sección 6 uno de los principales casos de uso de estos algoritmos es la clasificación de grandes cantidades de datos. Se propone en este estudio la clasificación de los eventos captados por el IACTs de CTAO según 'pedestal' o 'cosmic' a través de algoritmos de machine learning. Los eventos clasificados como 'flatfield' no se incluyen en la clasificación pues son eventos provocados intencionadamente con fines de calibración y están ya, por tanto, registrados y clasificados. Se eligieron para ello modelos simples de tres algoritmos distintos, K vecinos cercanos, vectores soporte y redes neuronales. Los modelos de K-vecinos cercanos y vectores soporte se implementaron con los parámetros casi por defecto de la librería scikit-learn [44] mientras que la red neuronal se creó a partir de la librería keras [45]. Cada modelo se entreno dos veces, la primera vez se hizo sobre los datos brutos de carga de phe por píxel y la segunda sobre el conjunto de nodos de una de las representaciones en grafo obtenidas. Por simplicidad de implementación técnica los grafos elegidos para entrenar los algoritmos de machine learning fueron los mostrados en la figura 13, es decir, aquellos con los parámetros de ajuste de ruido expresados como nodos independientes.

Como ya se vio en la sección 8 el problema de clasificación a tratar no está balanceado, es decir, hay disparidad entre el número total de los distintos tipos de eventos a clasificar; pedestal (744) y cosmic (39528). Este desbalance de clases puede derivar en el que al entrenar el modelo este no vea visto suficientes ejemplos de alguna de las clases como para aprender a reconocerla y que por tanto falle al clasificar nuevos datos de dicha clase. Existen diversas técnicas para afrontar esta situación, las más comúnmente usadas pasan por reducir el número de clases mayoritarias, aumentar de manera artificial el número de clases minoritarias o una combinación de ambas [46]. Dependiendo del campo de aplicación del algoritmo, estos métodos pueden presentar algunos problemas que, pueden llegar a desaconsejar su uso [47], [48], [49]. Forzar los algoritmos a obtener una buena puntuación sobre la predicción de cada clase mediante la introducción de datos artificiales conlleva a que estos solo aprendan de situaciones balanceadas e ideales que, por lo general, se alejan de la realidad sobre la que luego deberán predecir. Esta situación puede ser especialmente peligrosa en campos como la medicina dónde la salud de un paciente puede depender en parte de la capacidad de estos algoritmos. No obstante, dada la naturaleza del problema aquí tratado sí podría estar justificado replicar de manera artificial algunos de los eventos de 'pedestal' de forma que los modelos puedan clasificarlos con más facilidad. Esta justificación se aprecia de forma más clara a partir de la figura 9, debemos recordar que los datos de 'cosmic' y de 'ruido cosmic' pertenecen al mismo conjunto de datos, en la gráfica están separados solo con fines ilustrativos. Resulta claro entonces que los datos de 'pedestal' quedan por completo bajo la curva de los datos de 'cosmic', aumentar los datos de 'pedestal' de forma artificial podría ayudar a los modelos a reconocerlos. Los modelos se han evaluado mediante la matriz de confusión sobre

el conjunto de datos de testeo, dicha matriz representa de forma normalizada el número de falsos positivos y falsos negativos que el modelo ha predicho sobre el conjunto de testeo. Así, en caso de predecir todos los datos de manera precisa, la matriz será diagonal.

10.1 Limitaciones de los algoritmos de machine learning

La implementación de los distintos algoritmos de machine learning vino restringida por las capacidades de cada librería usada. Independientemente de las limitaciones para ajustar distintos parámetros, que no supusieron un impedimento pues no es el objetivo de este trabajo encontrar el mejor modelo. Los modelos de scikit-learn son modelos básicos, pensados principalmente para trabajar con datos bidimensionales. Inicialmente esto no supuso un problema pues el conjunto de datos inicial tiene dos dimensiones, el número de evento y los datos del mismo. Puesto que posteriormente los datos se expresan como grafos, el conjunto de datos adquiere una dimensión más, correspondiente a la unión entre nodos del grafo. Para poder estudiar si esta representación mediante grafos supone una mejora en la clasificación de los algoritmos se debería de entrenar los modelos antes usados sobre este nuevo conjunto de datos. Como se ha mencionado, los modelos de scikit-learn y su algoritmo para la sobrerepresentación de clases minoritarias aceptan solo datos bidimensionales, lo cual impide entrenar estos modelos sobre el nuevo conjunto de datos tridimensional lo que dificulta comparar que conjunto de datos se clasifica mejor. Se optó por tanto por crear un tercer conjunto de datos, nuevamente bidimensional formado a partir del conjunto de grafos, este nuevo conjunto se formó a partir del valor medio de cada conexión entre nodos. Se obtuvo así un conjunto que, pese a haber perdido la información relativa a la configuración del grafo, podía ser entrenado en los modelos ya creados y suponía una reinterpretación del conjunto original. Por su parte, si bien keras si permite crear modelos que acepten datos tridimensionales, el primer modelo fue diseñado para entrenarse sobre datos bidimensionales de manera que al obtener el dataset de grafos nos encontramos ante la misma problemática. Para poder realizar una comparativa se entrenó también este modelo sobre el tercer conjunto de datos creado a partir de la media de cada par de nodos. Así mismo, puesto que la librería lo permitía se creó otro modelo con una arquitectura diferente que aceptara datos tridimensionales. Al ser dos modelos con arquitecturas distintas, la comparación de sus resultados debería hacerse únicamente tras haberlos optimizado al máximo lo cual excede los propósitos de este trabajo. Es este último modelo, por tanto, el único que puede aprender sobre la configuración de nodos determinada por las hormigas, el resto de modelos verán versiones ligeramente modificadas del conjunto de datos original.

10.2 K-vecinos más cercanos

La implementación básica del algoritmo de los K-vecinos más cercanos de scikit-learn utiliza como métrica por defecto la euclídea y supone cinco vecinos. En el modelo propuesto se redujeron el número de vecinos a tres. Este modelo se

entreno sobre los datos en bruto, KNN-bruto, y sobre los datos una vez obtenidos los grafos, KNN-grafos, además se repitió cada entrenamiento sobre los mismos conjuntos de datos con la clase minoritaria sobrerrepresentada artificialmente, overKNN-grafos, overKNN-bruto.

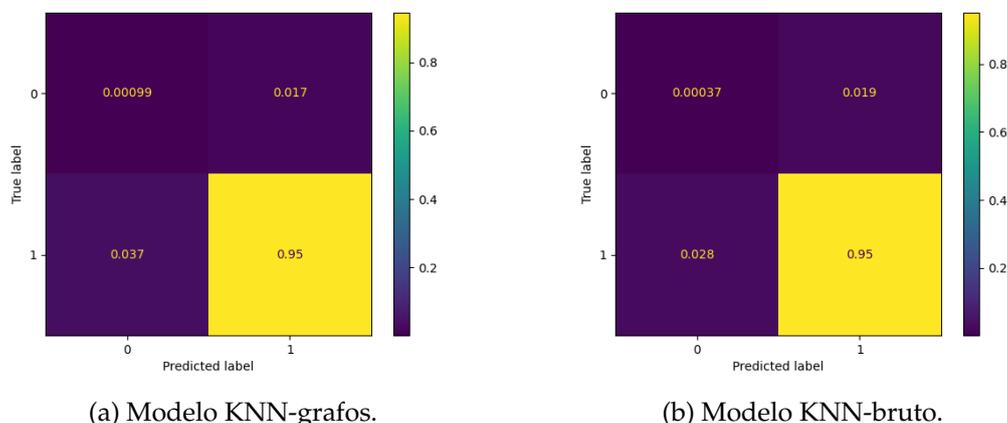


Figura 16: Matrices de confusión obtenidas del modelo de KNN entrenado con los datos en bruto y con los datos obtenidos de la representación de grafos. La etiqueta 0 corresponde a 'pedestal' la 1 corresponde a 'cosmic'.

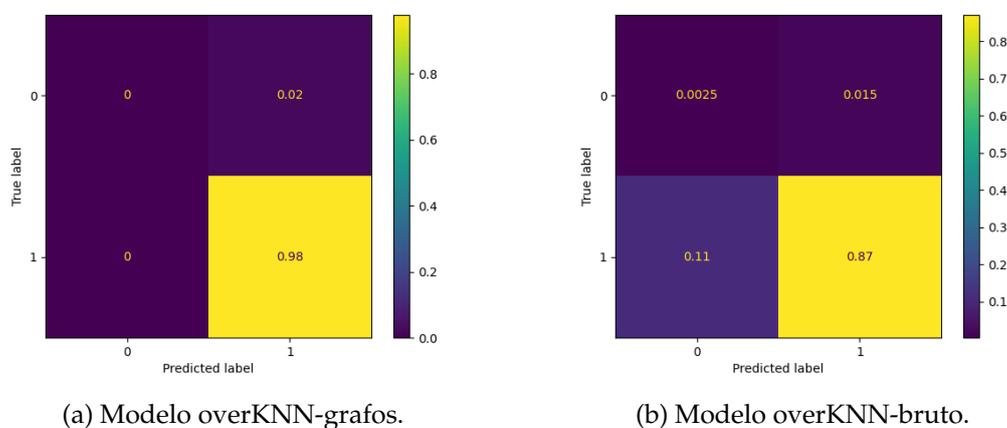


Figura 17: Matrices de confusión obtenidas del modelo de KNN entrenado con los datos en bruto y con los datos obtenidos de la representación de grafos con sobrerrepresentación de etiquetas minoritarias. La etiqueta 0 corresponde a 'pedestal' la 1 corresponde a 'cosmic'.

La sobrerrepresentación de los datos 'pedestal' supuso, en este caso, una mejora en la clasificación del modelo sobre los datos en bruto. Si bien ningún modelo consiguió clasificar correctamente más eventos 'pedestal' que los que clasificaba erróneamente, el modelo sobrerrepresentado de datos en bruto 17b fue el que mayor número de eventos 'pedestal' clasificó, a costa de la precisión, en la clasificación de eventos 'cosmic'. Por su parte, el modelo de grafos sobrerrepresentados 17a no clasificó correctamente ningún evento de 'pedestal'. Recordemos que este

modelo intenta clusterizar cada conjunto de datos de manera separada, en la representación por grafos todo el ruido se redujo a un solo conjunto de datos dado por la media y desviación del ruido, es decir, la clase 'pedestal' en este caso consta de un solo valor. Dicho valor se encuentra repetido incluso con mayor frecuencia en la clase 'cosmic'. Es por esto que en este caso, es más fácil clasificar de forma incorrecta las clases, esta dificultad pone de relieve los resultados obtenidos para el mismo conjunto de datos sin sobrerrepresentación 16a. Para el caso de los datos en bruto, como se ha visto, la sobrerrepresentación si puede suponer una mejora significativa en la clasificación pues el conjunto de valores de cada etiqueta será mayor y habrá por tanto más variedad de puntos que se podrán clasificar correctamente.

10.3 Vectores soporte, SVM

El algoritmo implementado por scikit-learn para el uso de SVM incluye diferentes opciones para personalizar el modelo según el problema a tratar, como se expuso en la sección 6 el kernel usado para nuestra clasificación fue el gaussiano. Para este kernel, scikit-learn define el parámetro C equivalente a λ en la expresión 6.2 y γ equivalente a $\frac{1}{\sigma}$ en la expresión 6.4. Recordamos que C regula cuanto se ajusta nuestra curva a los datos, es decir, cuantos datos clasificados erróneamente aceptamos, mientras que γ regula cuanta influencia tiene cada dato sobre sus vecinos. De igual modo que para el algoritmo de los k-vecinos cercanos se entrenaron cuatro modelos, dos para el conjunto de datos con etiquetas balanceadas, SVM-grafos y SVM-bruto, y dos para el conjunto balanceado, overSVM-grafos y overSVM-bruto.

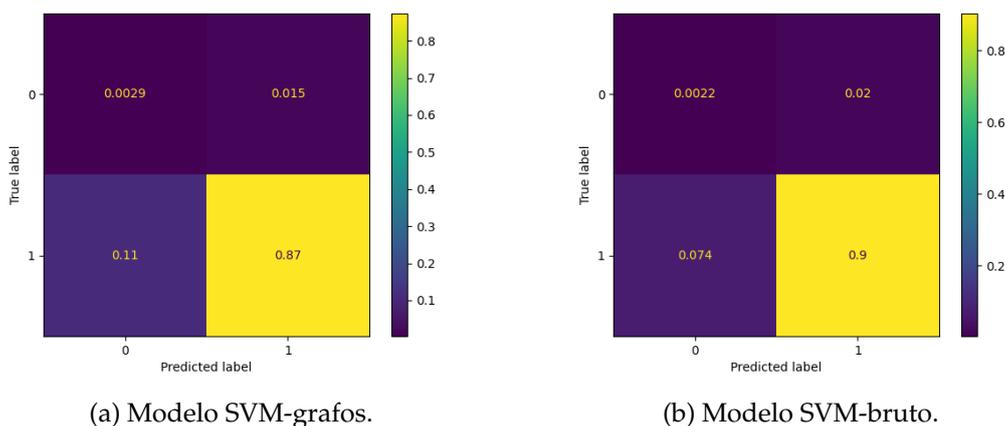


Figura 18: Matrices de confusión obtenidas del modelo de SVM entrenado con los datos en bruto y con los datos obtenidos de la representación de grafos. La etiqueta 0 corresponde a 'pedestal' la 1 corresponde a 'cosmic'.

Teóricamente, mediante el ajuste de sus distintos parámetros σ y λ , el algoritmo de SVM es capaz de clasificar correctamente cualquier problema mediante el kernel gaussiano pues en el caso más extremo se puede delinear cada conjunto de forma independiente mediante los parámetros reguladores. No obstante,

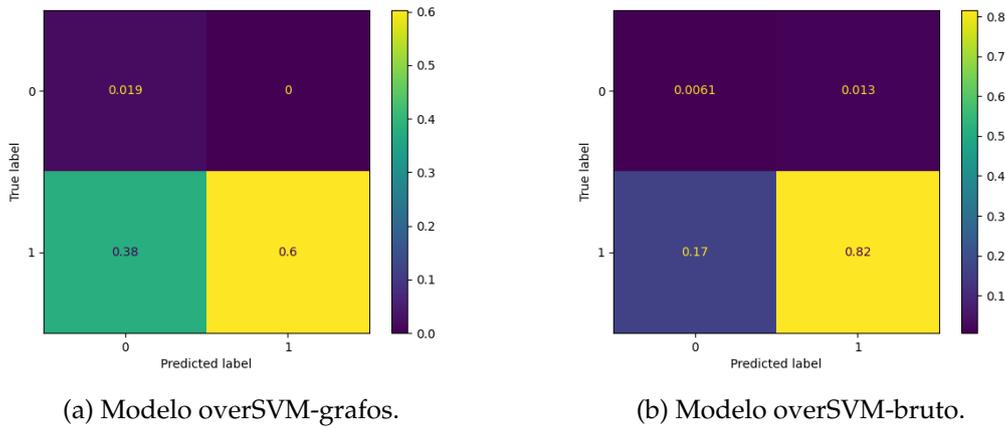


Figura 19: Matrices de confusión obtenidas del modelo de SVM entrenado con los datos en bruto y con los datos obtenidos de la representación de grafos con sobrerrepresentación de la etiqueta minoritaria. La etiqueta 0 corresponde a 'pedestal' la 1 corresponde a 'cosmic'.

la búsqueda de estos parámetros óptimos solo puede hacerse de forma manual, además, obtener una clasificación perfecta tampoco se consideraría un buen resultado pues implicaría que estamos sobreentrenando el modelo a nuestro conjunto de datos. Para los modelos aquí propuestos se ajustaron λ y σ teniendo en cuenta las características del conjunto de datos, para poder clasificar correctamente los eventos de 'pedestal' será necesario que la función kernel se ajuste bien al conjunto y deberemos ser laxos con las posibles muestras de 'cosmic' que se cuecen dentro del conjunto. Bajo este contexto, los mejores resultados se obtuvieron del modelo entrenado sobre el conjunto original de datos en bruto 18b, siendo incluso mejores que el modelo sobre datos en bruto sobrerrepresentado que en el caso de los K-vecinos dio los mejores resultados. Aunque nuevamente, ningún modelo consiguió clasificar los eventos 'pedestal' correctamente de forma mayoritaria, en este caso el conjunto de datos sobrerrepresentado no supuso una mejoría en ninguno de los casos respecto a los conjuntos originales. Para el caso de los conjuntos de grafos, al estar todo el ruido localizado en un punto, al sobrerrepresentarlo el modelo se sobreentrenó en esta clase 19a, etiquetando erróneamente los datos de 'cosmic'.

10.4 Red neuronal

Para el entrenamiento de la red neuronal no se usó la librería scikit-learn pues su simplicidad de implementación restringe en gran medida las posibilidades de personalización de la red. Se optó, en cambio, por crear la red capa a capa mediante la librería keras. Esta librería permitió la creación de dos modelos, el primero de 18571 parámetros se entrenó sobre los datos en 2 dimensiones, tanto brutos como en grafos como sobrerrepresentados, cuatro entrenamientos en total, NN2D-grafos, NN2D-bruto, overNN2D-grafos, overNN2D-bruto. El segundo de apenas 1320 parámetros NN-3D, en cambio, se diseñó con una arquitectura distinta con el propósito de que pudiera extraer información sobre el conjunto de datos de los

grafos en su versión tridimensional, es decir, se pretende mediante la implementación de una capa convolucional que este modelo aprenda no solo sobre los datos en si no también sobre el modo en el que las hormigas los han interconectado. En ambos modelos, excepto en los casos con sobrerrepresentación, se realizó un ligero ajuste del peso de cada etiqueta con el fin de que el modelo pudiera clasificar mejor los datos.

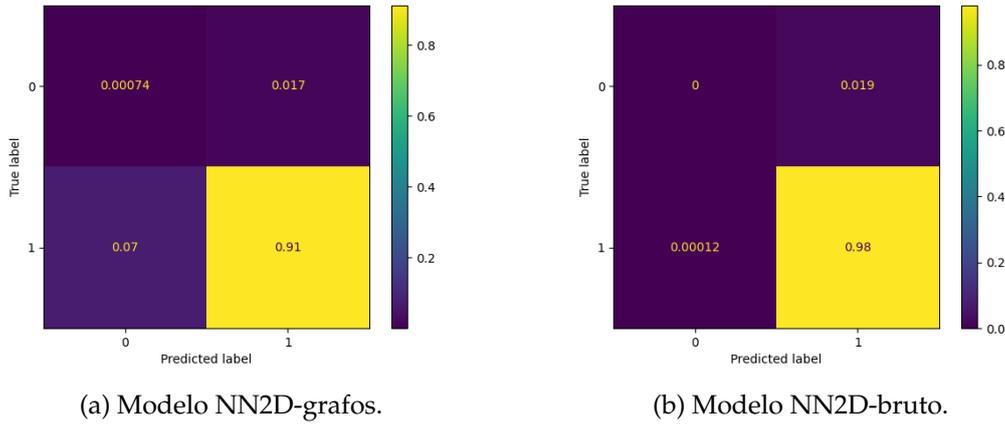


Figura 20: Matrices de confusión obtenidas del modelo de NN2D entrenado con los datos en bruto y con los datos obtenidos de la representación de grafos. La etiqueta 0 corresponde a 'pedestal' la 1 corresponde a 'cosmic'.

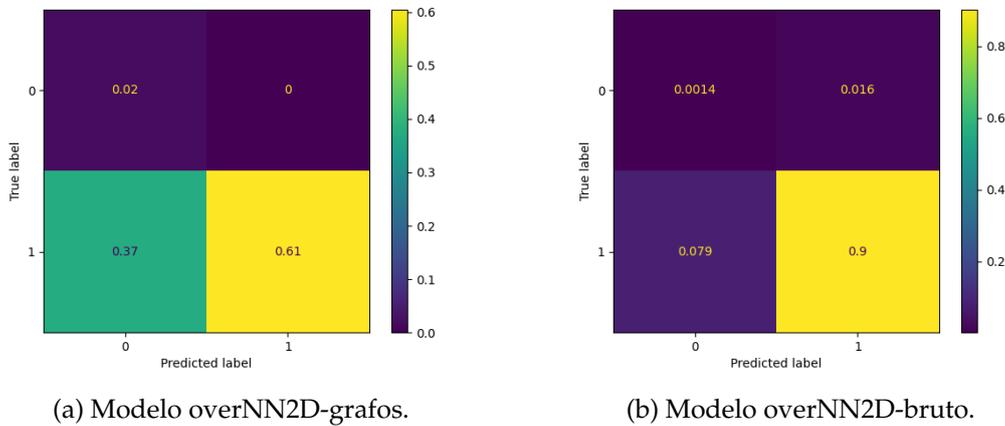


Figura 21: Matrices de confusión obtenidas del modelo de NN2D entrenado con los datos en bruto y con los datos obtenidos de la representación de grafos con sobrerrepresentación de etiquetas minoritarias. La etiqueta 0 corresponde a 'pedestal' la 1 corresponde a 'cosmic'.

Al igual que se observó en el caso de los K-vecinos cercanos, sobrerrepresentar la clase minoritaria supuso una mejora de la clasificación del conjunto de testeo respecto al resto de modelos de NN e incluso mejor que el modelo homogéneo de KNN [21b](#). Por su parte, nuevamente, la sobrerrepresentación sobre el conjunto de grafos [21a](#) supuso una peor clasificación, puesto que al expresar todo el

ruido como el mismo dato el modelo se sobreentrenó sobre este. En este caso la representación mediante grafos desbalanceada sí mostró mejores resultados pues el modelo, en comparación con los resultados sobre los datos en bruto, empezó a reconocer los eventos de 'pedestal' mientras seguía clasificando correctamente los 'cosmic'. No obstante, debido a la naturaleza de los datos este modelo de red neuronal no aprendió nada sobre la configuración de datos elegida por las hormigas. Al sobrerrepresentar los datos en bruto, al tener más variedad de la clase infrarepresentada, el modelo aprendió a distinguir los datos de 'pedestal' mejor que en el conjunto de datos en bruto desequilibrado [20b](#).

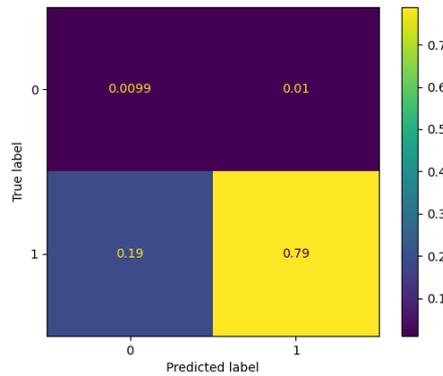


Figura 22: Matriz de confusión obtenida del modelo de NN3D con los datos tridimensionales. La etiqueta 0 corresponde a 'pedestal' la 1 corresponde a 'cosmic'.

Pese a la diferencia de parámetros respecto al modelo para datos bidimensionales, el modelo NN-3D, mostró resultados aceptables, clasificando mejor los eventos 'pedestal' que la mayoría de modelos ya mostrados, aunque, a cambio mostró menos precisión al clasificar los eventos de 'cosmic' [22](#). No obstante, estos resultados sugieren que sería posible obtener una clasificación mejor mediante un ajuste adecuado de los parámetros de entrenamiento o incluso de la arquitectura del propio modelo. Esto podría suponer añadir más neuronas, más capas, o incluso cambiar el concepto y usar otro tipo de capas en lugar de las convolucionales como las LSTM (Long Short Term Memory) las cuales deberían ser capaces de retener la información relativa a la configuración de nodos.

11 Conclusiones

Se ha estudiado la posibilidad de una representación alternativa de los datos de carga por píxel obtenidos por el primer prototipo de telescopio de imagen atmosférica Cherenkov de CTAO, comparando además el espacio en memoria ocupado por ambas representaciones. Así mismo, se han considerado diversas técnicas de clasificación de los datos en sus dos versiones. Dicha representación fue obtenida expresando los datos como grafos de los cuales posteriormente, mediante el algoritmo de las hormigas, se encontró de forma heurística la representación óptima. Por su parte, la clasificación se realizó mediante algoritmos de machine learning

como k-vecinos, vectores soporte y redes neuronales para ambos, conjunto de datos original y el obtenido a partir de la representación por grafos. El algoritmo de las hormigas se implementó en su versión más básica, así como los parámetros que lo configuraban. Por defecto, cada hormiga empieza su recorrido desde nodos distintos, esto dificultó, comprobar si partiendo del mismo nodo las hormigas convergían al mismo camino al correr el algoritmo varias veces. Al seleccionar un número de hormigas y de iteraciones elevado para suplir dicha aleatoriedad inicial las hormigas tampoco convergieron al mismo camino. Esto puede deberse al resto de parámetros que influyen en el algoritmo como el peso o nivel de feromona inicial de cada nodo, la importancia de la distancia entre nodos o la importancia de su peso, los cuales no se modificaron y se configuraron por defecto. No obstante, para el caso de estudio elegido, hay algunos patrones que si parecen repetirse como que los nodos pertenecientes al ruido siempre aparecen juntos o que los nodos de mayor energía no se encuentran seguidos. Confirmar estas observaciones, en cambio, requeriría un estudio más dedicado sobre varios eventos teniendo en cuenta los distintos parámetros mencionados. Por otra parte, generar el conjunto de datos de todos los grafos supuso varios días de computación continuada, el cual es uno de los motivos por los que se prevé necesario explorar otras configuraciones de los datos originales para crear los grafos además de los propios parámetros de las hormigas. Otro de los motivos por los que explorar otras configuraciones se presenta necesario son los resultados de los algoritmos de clasificación. Si bien es cierto que el único modelo que pudo clasificar los datos de los grafos como tal y por tanto aprender sobre las conexiones entre nodos fue el algoritmo de red convolucional, los resultados obtenidos no supusieron una mejora con respecto al resto de modelos que clasificaban sobre los datos originales. Por su parte, la reducción del conjunto de datos de grafos tridimensional al bidimensional llevó a los mismos resultados que el conjunto de datos bruto, lo cual se explica por la similitud entre ambos conjuntos. Al hacer la media entre cada par de nodos los datos obtenidos no distan de los originales lo suficiente como para suponer una diferencia al clasificarlos. Una representación alternativa de los datos, quizás incluyendo la información temporal, podría facilitar su clasificación, así como un conjunto mayor de datos que pudiera mostrar más variedad de cada evento. Otras arquitecturas, como LSTM, redes de clasificación de imágenes, transformers... , sobre todo para los datos tridimensionales podrían dar también lugar a resultados diferentes. Pese a que no hubo una mejora en la clasificación la reducción del espacio en memoria sí resultó ser significativa lo cual, dada la novedad de esta metodología, supone un importante punto de partida para futuras investigaciones en esta línea. Estudios futuros deberán determinar si a partir de esta u otra representación de los grafos es posible determinar si el evento observado corresponde a un rayo gamma o un rayo cósmico y en caso de no ser así estudiar si la reducción en el espacio de almacenamiento compensa esta posible pérdida de información.

Referencias

- ¹C. Council, *Cherenkov Telescope Array Observatory*, <https://www.cta-observatory.org/>.
- ²E. Oks, "Brief review of recent advances in understanding dark matter and dark energy", *New Astronomy Reviews* **93**, 101632 (2021).
- ³B. Carr y F. Kühnel, "Primordial black holes as dark matter: recent developments", *Annual Review of Nuclear and Particle Science* **70**, 355-394 (2020).
- ⁴L. Boyle, K. Finn y N. Turok, "CPT -Symmetric Universe", *Physical Review Letters* **121**, Cited by: 77; All Open Access, Green Open Access, Hybrid Gold Open Access, 10.1103/PhysRevLett.121.251301 (2018).
- ⁵M. Bashkanov y D. Watts, "A new possibility for light-quark dark matter", *Journal of Physics G: Nuclear and Particle Physics* **47**, 03LT01 (2020).
- ⁶R. A. Millikan, *Award of the Nobel Prize in Physics to Victor F. Hess and Carl D. Anderson*, 1937.
- ⁷W. Davis, "Discovery of the Positron", *Science* **84**, 8-9 (1936).
- ⁸T. Lee, "A brief history of the muon", *Hyperfine Interactions* **86**, 439-453 (1994).
- ⁹D. Perkins, "The Discovery of the Pion in Bristol in 1947", en *Heavy Flavour Physics: a Probe of Nature's Grand Design* (IOS Press, 1998), págs. 1-12.
- ¹⁰R. H. Dalitz, "Kaon physics—The first 50+ years", *Kaon Physics*, 1 (2001).
- ¹¹T. Nakano y K. Nishijima, "Charge Independence for V-particles", *Progress of Theoretical Physics* **10**, 581-582 (1953).
- ¹²A. Bell, "The acceleration of cosmic rays in shock fronts—II", *Monthly Notices of the Royal Astronomical Society* **182**, 443-455 (1978).
- ¹³Y. Uchiyama, F. A. Aharonian, T. Tanaka, T. Takahashi e Y. Maeda, "Extremely fast acceleration of cosmic rays in a supernova remnant", *Nature* **449**, 576-578 (2007).
- ¹⁴I. Cholis, L. Goodenough, D. Hooper, M. Simet y N. Weiner, "High energy positrons from annihilating dark matter", *Physical Review D* **80**, 123511 (2009).
- ¹⁵Q. Yuan y L. Feng, "Dark Matter Particle Explorer observations of high-energy cosmic ray electrons plus positrons and their physical implications", *SCIENCE CHINA Physics, Mechanics & Astronomy* **61**, 1-12 (2018).
- ¹⁶E. Aliu, H. Anderhub, L. t. Antonelli, P. Antoranz, M. Backes, C. Baixeras, J. Barrio, H. Bartko, D. Bastieri, J. Becker et al., "Observation of pulsed γ -Rays above 25 GeV from the Crab Pulsar with MAGIC", *Science* **322**, 1221-1224 (2008).
- ¹⁷V. Acciari, E. Aliu, T. Arlen, T. Aune, M. Beilicke, W. Benbow, S. Bradbury, J. Buckley, V. Bugaev, K. Byrum et al., "VERITAS observations of gamma-ray bursts detected by Swift", *The Astrophysical Journal* **743**, 62 (2011).
- ¹⁸J. Aleksić, E. Alvarez, L. Antonelli, P. Antoranz, M. Asensio, M. Backes, J. Barrio, D. Bastieri, J. B. González, W. Bednarek et al., "Phase-resolved energy spectra of the Crab pulsar in the range of 50–400 GeV measured with the MAGIC telescopes", *Astronomy & astrophysics* **540**, A69 (2012).

- ¹⁹M. Collaboration, V. Acciari, S. Ansoldi, L. Antonelli, A. Arbet Engels, C. Arcaro, D. Baack, A. Babić, B. Banerjee, P. Bangale et al., “Discovery of TeV γ -ray emission from the neighbourhood of the supernova remnant G24. 7+ 0.6 by MAGIC”, *Monthly Notices of the Royal Astronomical Society* **483**, 4578-4585 (2019).
- ²⁰V. A. Acciari, S. Ansoldi, L. Antonelli, A. A. Engels, K. Asano, D. Baack, A. Babić, A. Baquero, U. B. De Almeida, J. Barrio et al., “Detection of the Geminga pulsar with MAGIC hints at a power-law tail emission beyond 15 GeV”, *Astronomy & astrophysics* **643**, L14 (2020).
- ²¹H. Abdalla, A. Abramowski, F. Aharonian, F. A. Benkhali, E. Angüner, M. Arakawa, M. Arrieta, P. Aubert, M. Backes, A. Balzer et al., “The HESS Galactic plane survey”, *Astronomy & Astrophysics* **612**, A1 (2018).
- ²²S. Ahmed, “Scintillation detectors and photodetectors”, *Physics and Engineering of Radiation Detection (Second Edition)*. Ed. by SN Ahmed. Elsevier, 331-433 (2015).
- ²³A. M. Hillas, “Cherenkov light images of EAS produced by primary gamma”, en 19th Intern. Cosmic Ray Conf-Vol. 3, OG-9.5-3 (1985).
- ²⁴J. Albert, E. Aliu, H. Anderhub, P. Antoranz, A. Armada, M. Asensio, C. Baixeras, J. Barrio, H. Bartko, D. Bastieri et al., “Implementation of the random forest method for the imaging atmospheric Cherenkov telescope MAGIC”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **588**, 424-432 (2008).
- ²⁵D. Nieto, A. Brill, B. Kim y T. Humensky, “Exploring deep learning as an event classification method for the Cherenkov Telescope Array”, arXiv preprint arXiv:1709.05889 (2017).
- ²⁶M. De Naurois, J. Guy, A. Djannati-Atai y J. Tavernet, “Application of an analysis method based on a semi-analytical shower model to the first HESS telescope”, en *International Cosmic Ray Conference 28 ICRC 2003* (Universal Academy Press, 2003), págs. 2907-2910.
- ²⁷J. Cortina y C. Delgado, “Detecting gamma rays with high resolution and moderate field of view: the air Cherenkov technique”, en *Handbook of X-ray and Gamma-ray Astrophysics* (Springer, 2024), págs. 2547-2573.
- ²⁸V. A. Acciari, S. Ansoldi, L. A. Antonelli, A. A. Engels, K. Asano, D. Baack, A. Babić, A. Baquero, U. B. de Almeida, J. Barrio et al., “MAGIC observations of the nearby short gamma-ray burst GRB 160821B”, *The Astrophysical Journal* **908**, 90 (2021).
- ²⁹R. López-Coto, A. Moralejo, M. Artero, A. Baquero, M. Bernardos, J. L. Contreras, F. Di Pierro, E. García, D. Kerszberg, M. López-Moya, A. MasAguilar, D. Morcuende, M. Noethe, S. Nozaki, Y. Ohtani, C. Priyadarshi, Y. Suda, T. Vuillaume et al., “Physics Performance of the Large Size Telescope prototype of the Cherenkov Telescope Array”, en *Proceedings, 37th International Cosmic Ray Conference*, vol. 395 (2021), pág. 806.
- ³⁰N. Monmarché, F. Guinand y P. Siarry, *Artificial Ants* (Wiley, 2010).
- ³¹M. Dorigo y L. Gambardella, “Ant colony system: a cooperative learning approach to the traveling salesman problem”, *IEEE Transactions on Evolutionary Computation* **1**, 53-66 (1997).
- ³²rhgrant10, *Ant Colony Optimization for Tthe Traveling Salesman Problem*. <https://pypi.org/project/acopy/>.

- ³³J. Byers, "The physics of data", *Nature Physics* **13**, 10.1038/nphys4202 (2017).
- ³⁴G. Longo, E. Merényi y P. Tiño, "Foreword to the focus issue on machine intelligence in astronomy and astrophysics", *Publications of the Astronomical Society of the Pacific* **131**, 1-6 (2019).
- ³⁵J.-V. Rodríguez, I. Rodríguez-Rodríguez y W. L. Woo, "On the application of machine learning in astronomy and astrophysics: A text-mining-based scientometric analysis", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **12**, e1476 (2022).
- ³⁶A. Tanaka y A. Tomiya, "Detection of phase transition via convolutional neural networks", *Journal of the Physical Society of Japan* **86**, 063001 (2017).
- ³⁷K. Shinjo, S. Sota, S. Yunoki y T. Tohyama, "Characterization of photoexcited states in the half-filled one-dimensional extended Hubbard model assisted by machine learning", *Physical Review B* **101**, 195136 (2020).
- ³⁸P. Baireuther, T. E. O'Brien, B. Tarasinski y C. W. Beenakker, "Machine-learning-assisted correction of correlated qubit errors in a topological code", *Quantum* **2**, 48 (2018).
- ³⁹P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher y D. J. Schwab, "A high-bias, low-variance introduction to machine learning for physicists", *Physics reports* **810**, 1-124 (2019).
- ⁴⁰J. Fuchs, S. Goldt y U. Seifert, "Stochastic thermodynamics of resetting", *Europhysics Letters* **113**, 60009 (2016).
- ⁴¹D. P. Kingma y J. Ba, "Adam: A method for stochastic optimization", *arXiv preprint arXiv:1412.6980* (2014).
- ⁴²A. A. Abud, B. Abi, R. Acciarri, M. Acero, M. Adames, G. Adamov, M. Adamowski, D. Adams, M. Adinolfi, C. Adriano et al., "Reconstruction of interactions in the ProtoDUNE-SP detector with Pandora", *The European Physical Journal C* **83**, 618 (2023).
- ⁴³A. A. Hagberg, D. A. Schult y P. J. Swart, "Exploring Network Structure, Dynamics, and Function using NetworkX", en *Proceedings of the 7th Python in Science Conference*, ed. por G. Varoquaux, T. Vaught y J. Millman (2008), págs. 11-15.
- ⁴⁴F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot y E. Duchesnay, "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research* **12**, 2825-2830 (2011).
- ⁴⁵F. Chollet et al., *Keras*, <https://keras.io>, 2015.
- ⁴⁶N. V. Chawla, K. W. Bowyer, L. O. Hall y W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique", *Journal of artificial intelligence research* **16**, 321-357 (2002).
- ⁴⁷A. B. Hassanat, A. S. Tarawneh, G. A. Altarawneh y A. Almuhaimeed, "Stop oversampling for class imbalance learning: A critical review", *arXiv preprint arXiv:2202.03579* (2022).
- ⁴⁸A. Carriero, K. Luijken, A. de Hond, K. G. Moons, B. van Calster y M. van Smeden, *The harms of class imbalance corrections for machine learning based prediction models: a simulation study*, 2024.

-
- ⁴⁹R. van den Goorbergh, M. van Smeden, D. Timmerman y B. Van Calster, "The harm of class imbalance corrections for risk prediction models: illustration and simulation using logistic regression", *Journal of the American Medical Informatics Association* **29**, 1525-1534 (2022).



Propuesta de Trabajo Fin de Grado en Física

Tutor/a:	<i>Alberto Guillén Perales</i>
Departamento y Área de Conocimiento:	<i>Ingeniería de Computadores, Automática y Robótica/ATC</i>
Correo electrónico:	<i>aguillen@ugr.es</i>
Cotutor/a:	<i>Rubén López Coto</i>
Departamento y Área de Conocimiento:	<i>IAA</i>
Correo electrónico:	<i>rlopezcoto@iaa.es</i>

Título del Trabajo: Procesado de imágenes de CTA mediante algoritmos de hormigas			
Tipología del Trabajo: (Según punto 3 de las Directrices del TFG aprobadas por Comisión Docente el 10/12/14)	(Marcar con X)	1. Revisión bibliográfica	
		2. Estudio de casos teórico-prácticos	
		3. Trabajos experimentales	X
		4. Elaboración de nuevas prácticas de laboratorio	
		5. Elaboración de un proyecto	
		6. Trabajo relacionado con prácticas externas	

Breve descripción del trabajo:

Se propone procesar las imágenes obtenidas en el observatorio CTA para transformar esa información mediante la salida de un algoritmo basado en el comportamiento de las colonias de hormigas

Objetivos planteados:

Implementar un algoritmo de colonia de hormigas o implementar el uso de una biblioteca
Aplicar dicho algoritmo a las imágenes del observatorio
Analizar los resultados obtenidos comparando el potencial ahorro de almacenamiento y cómputo

Metodología:

Resolver un problema básico con ACO
EDA de las imágenes de CTA
Aplicar ACO a imágenes de CTA

Bibliografía:

<https://pypi.org/project/aco/>
<https://www.cta-observatory.org/>
Monmarché Nicolas, Guinand Frédéric and Siarry Patrick (2010). *Artificial Ants*. Wiley-ISTE. ISBN .
Dorigo, Gambardella, M, L.M. (1997). "Learning Approach to the Traveling Salesman Problem". *IEEE Transactions on Evolutionary Computation*. 1 (1): 214. doi:10.1109/4235.585892.

A rellenar sólo en el caso que el alumno sea quien realice la propuesta de TFG



UNIVERSIDAD
DE GRANADA



Facultad de Ciencias
Sección de Físicas

Alumno/a propuesto/a:

Granada, 24 de

Mayo 2023

Sello del Departamento