

Universidad de Oviedo

DESARROLLO DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA LA IDENTIFICACIÓN DE LEPTONES EN EL EXPERIMENTO CMS DEL CERN

Sara Iglesias Suárez-Noguerol

Dirigido por Bárbara Álvarez González y Clara Ramón Álvarez

> UNIVERSIDAD DE OVIEDO Facultad de Ciencias Grado en Física

> > Julio de 2022

Índice general

1.	Con	ceptos Básicos	7
	1.1.	Modelo estándar	7
		1.1.1. Interacciones fundamentales y partículas portadoras de las interacciones	8
		1.1.2. Partículas constituyentes de la materia	9
2.	Disj	positivo experimental	12
	2.1.	El LHC	12
	2.2.	Magnitudes relevantes	14
	2.3.	CMS	15
		2.3.1. Sistema de detección de trazas	16
		2.3.2. Calorímetro electromagnético	17
		2.3.3. Calorímetro de hadrones	17

	2.3.4. Sistema de Muones	18
3. Ree	construcción de objetos en CMS	20
3.1.	Particle Flow	21
	3.1.1. Reconstrucción local	21
	3.1.2. Algoritmo de unión (Link)	22
4. Pro	ocesos Físicos de Estudio	27
4.1.	Muones que provienen de bososnes	28
	4.1.1. Proceso Drell-Yan	29
	4.1.2. Procesos $t\bar{t}$	30
	4.1.3. Producción y desintegración del bosón de Higgs	31
4.2.	Muones que provienen de mesones	33
	4.2.1. Desintegración de hadrones B	34
5. Int:	roducción al Machine Learning	36
5.1.	Sistemas de Machine Learning	37
	5.1.1. Aprendizaje supervisado vs no supervisado	37
	5.1.2. Aprendizaje online vs tipo <i>batch</i>	39

		5.1.3.	Aprendizaje basado en instancias vs basado en modelos	39
	5.2.	Modelo	os usados en este trabajo	40
		5.2.1.	Árboles de decisión	43
		5.2.2.	Random Forest	44
		5.2.3.	Boosted Decision Trees	45
6.	Aná	ilisis m	ultivariante para la clasificación muones	47
	6.1.	Simula	ciones de Monte Carlo	47
		6.1.1.	Muestra empleada en este trabajo	49
	6.2.	Variabl	les de entrada del MVA	51
		6.2.1.	Momento transverso p_t y pseudorapide z η	55
		6.2.2.	Parámetros de impacto, dxy y dz	57
	6.3.	Variab	le de salida del MVA	58
7.	Moo	delos ei	ntrenados	61
	7.1.	Rando	m Forest sin dxy y dz y señal $p_T > 10 \; GeV \; \ldots \; \ldots$	62
	7.2.	Randor	m Forest con todas las variables y señal $p_T > 10~GeV$	66

7.3.	Estudio de la distribución de los parámetros de impacto dxy y dz	71
7.4.	Random Forest sin dxy y dz y señal $p_T > 20 \ GeV$	73
7.5.	Random Forest sin dxy y dz,p_T y $\eta,$ y señal $p_T>10\;GeV$.	76
7.6.	Boosted Decision Trees sin dxy y dz y señal $p_T>10\;GeV$.	79
7.7.	Comparación de los modelos	83

Introducción

El LHC es el acelerador de partículas que alcanza una mayor energía hasta la fecha. Este acelerador colisiona protones e iones pesados a energías del orden de TeV. Proporciona, por tanto, unas condiciones únicas para poner a prueba el modelo estándar de la física de partículas, realizar búsquedas de partículas supersimétricas, materia oscura, etc. Por ser éste un colisionador hadrónico, los estados finales con hadrones serán dominantes. Por lo tanto, los procesos que contengan leptones en el estado final son especialmente interesantes, pues la presencia de estas partículas puede ser indicativo de que el proceso físico que ha tenido lugar en la colisión era relevante.

El LHC genera una cantidad muy alta de datos durante sus periodosactivos. Las colisiones que tienen lugar en el LHC son registradas en forma de señales eléctricas por los experimentos que se encuentran distribuidos en el acelerador. En cada experimento, el detector debe ser capaz de combinar todas esas señales eléctricas para determinar qué partículas se han generado en el estado final y qué características tienen. Este proceso de reconstrucción e identificación es complejo y de vital importancia para el posterior análisis físico de los datos. Así, es imprescindible la continua mejora de las técnicas de procesamiento para tal volumen de datos, de forma que se optimice el coste computacional y se reduzca en lo posible el error de identificación.

Durante el periodo de toma de datos comprendido entre 2015 y 2018 (llamado Run 2), la identificación del muones (tipo de leptón que se va a usar para los estudios de este trabajo) fue muy buena, presentando una eficiencia superior al 85 %. No obstante, con la perspectiva del inicio del Run 3 en julio de 2022 y, teniendo en cuenta que se espera recoger una cantidad de datos todavía mayor a la del Run 2 es necesario intentar mejorar la identificación

de muones.

Por ello en este trabajo se estudiará la aplicación de diferentes técnicas de inteligencia artificial a la identificación de muones a partir de los datos registrados por el el experimento CMS del LHC en colisiones protón-protón a 13 TeV durante el año 2018.

El objetivo principal será diseñar el modelo de aprendizaje automático, en particular un modelo de clasificación, que mejor seleccione los muones. Para ello, se escogerá tanto el algoritmo y sus hiperparámetros, como las características físicas que este puede tener en cuenta para la clasificación.

Este trabajo se enmarca en el contexto de las siguientes asignaturas del grado: Física Nuclear y de Partículas Elementales, Métodos Numéricos Aplicados a la Física y Técnicas Experimentales III.

La memoria se estructura en distintos bloques: primero, se introducen conceptos básicos para la física de partículas y el modelo estándar (capítulo 1). A continuación, se presenta el dispositivo experimental que se empleará en este trabajo, que son el acelerador LHC y el detector CMS, (capítulo 2) y se describe el proceso de reconstrucción de las partículas (capítulo 3). En el capítulo 4 se exponen algunos de los procesos físicos cuyo estudio se beneficiará de la buena identificación de muones. En el capítulo 5 se introducen los conceptos básicos del aprendizaje automático y los algoritmos que se usarán en este estudio. En el capítulo 6 se presenta el planteamiento del problema de identificación de muones a resolver y en el capítulo 7 se detallan todos los modelos entrenados. Finalmente se recogen las conclusiones obtenidas, así como posibles líneas futuras de investigación.

Capítulo 1

Conceptos Básicos

1.1. Modelo estándar

El modelo estándar (ME) es una teoría de la física de partículas desarrollada a partir de 1970 que engloba lo que se conoce hasta hoy sobre la estructura de la materia. Según el ME, el universo está formado por partículas elementales, fermiones, y bosones. Estos últimos son portadores de tres fuerzas fundamentales.

El ME es una teoría robusta que hasta el momento ha conseguido superar todas las pruebas experimentales realizadas en su contexto. A pesar de ello, tiene ciertas limitaciones, pues no es capaz de explicar, por ejemplo, la interacción gravitatoria o la composición de la materia oscura.

Lo expuesto en este capítulo ha sido extraído tanto del libro *Particles and nucleai* de Bogdan Povh [14], como del artículo *The Standard Model* en la página web del CERN [17].

1.1.1. Interacciones fundamentales y partículas portadoras de las interacciones

En el ME se recogen tres interacciones fundamentales, cada una de ellas con una intensidad y un rango de interacción diferente. Aunque no se recoge en el ME, existe una cuarta interacción fundamental, la gravitacional:

- La interacción *gravitacional*: tiene rango infinito, es la de intensidad más débil (relativamente, de orden 1). La experimentan todas las partículas con masa y sólo tiene carácter atractivo.
- La interacción *electromagnética*: tiene rango infinito, pero una intensidad relativa del orden 10³⁶. Actúa sobre las partículas cargadas eléctricamente.
- La interacción fuerte: esta interacción tiene un rango de tan sólo $10^{-15} m$, pero una fuerza relativa 10^{38} veces mayor que la gravitacional. Las partículas sobre las que actúa son aquellas que tienen carga de color.
- La interacción $d\acute{e}bil$: es la de menor radio de interacción, de $10^{-18} m$, con una fuerza relativa del orden de 10^{25} . La experimentan partículas con carga débil.

La fuerza gravitatoria es insignificante a nivel de la física subatómica, debido a su baja intensidad no influye en la interacción de partículas elementales. Para las tres restantes, se recoge en el modelo estándar que están mediadas por el intercambio de *bosones vectoriales*, partículas portadoras de interacción con espín 1. Estas serán:

- El *gluon* en caso de la fuerza fuerte. Es una partícula de masa y carga eléctrica nulas, pero con carga de color, es decir, también experimenta la interacción fuerte.
- El *fotón* en el caso de la electromagnética. El fotón tiene también masa y carga eléctrica nula y es su propia antipartícula.
- Los bosones W^+ , W^- y Z^0 , para la interacción débil. No tienen carga de color (no interactúan fuertemente), su carga eléctrica es +1, -1 y

0, respectivamente. Son partículas con masa, lo que causa el pequeño rango de interacción de la fuerza débil. El bosón W^+ es la antipartícula del W^- , y el Z^0 es su propia antipartícula.

Por el momento, a pesar de que se ha teorizado sobre ella, la partícula portadora de la interacción gravitatoria, el *gravitón*, no se ha encontrado experimentalmente.

Finalmente, se incluye en el modelo estándar un bosón escalar, el bosón de Higgs, sin carga eléctrica ni de color. Esta partícula interacciona tanto con fermiones como con los bosones W^{\pm} y Z^{0} , dándoles masa mediante el mecanismo de Higgs.

1.1.2. Partículas constituyentes de la materia

Los fermiones, según el ME, se dividen en dos categorías: leptones y quarks. Cada grupo consta de seis partículas, clasificadas en familias o *generaciones* en las que se comparten diversas propiedades. Tanto los leptones como los quarks tienen espín 1/2. Al contrario que para átomos, núcleos atómicos o hadrones, no se han observado estados excitados para ninguno de los fermiones.

Las partículas más ligeras y estables forman la primera generación, mientras que las más pesadas e inestables constituyen la segunda y tercera. Así, las partículas de la primera generación conforman toda la materia estable del universo, mientras que el resto se desintegra en partículas más estables.

La primera generación de leptones está compuesta por el electrón e y el neutrino de electrón ν_e , la segunda por el muon μ y el neutrino del muon ν_{μ} y en la tercera, el tau τ y el neutrino del tau ν_{τ} . El electrón, muon y tau tienen todos carga eléctrica y masa considerable (~ 0.5 MeV/c^2 , 10³ MeV/c^2 y 10³ MeV/c^2 , respectivamente), mientras que los neutrinos son eléctricamente neutros y, aunque se sabe que su masa es muy baja, esta no se ha medido todavía.

Los seis quarks están emparejados en tres generaciones: el quark up y el down forman la primera generación, el quark *charm* y el quark *strange* la segunda y el quark *top* y *bottom* o *beauty* forman la tercera.

Una propiedad específica de los quarks es que cuentan con carga de color y, en la naturaleza, no pueden aparecer aislados, si no que se combinan de forma que esta sea nula. Esta propiedad se llama confinamiento cuántico y se debe a la interacción de los gluones consigo mismos. Las partículas formadas por quarks reciben el nombre de *hadrones*, y permanecen unidos debido a la interacción fuerte entre ellos.

Como los gluones y quarks no pueden aparecer aislados, cuando estos se producen de forma solitaria tras una colisión, se unen a otros quarks del vacío cuántico para formar hadrones. Se dice en estos casos que el gluon o quark solitario se **hadroniza**, y las agrupaciones de hadrones que se observan como resultado se conocen como *chorros de partículas hadrónicas* o *jets hadrónicos*.

En la figura 1.1 se resume lo expuesto en esta sección.



Figura 1.1: El modelo de partículas de partículas elementales. Las tres primeras columnas corresponden con las tres generaciones de fermiones, respectivamente: en morado los quarks, en verde los leptones. La cuarta columna corresponde a los bosones vectoriales, partículas portadoras de interacción. Por último, se incluye el bosón escalar de Higgs. [18]

Capítulo 2

Dispositivo experimental

2.1. El LHC

El LHC (*Large Hadron Collider*) es el acelerador de partículas más grande y potente del mundo, y la última adición al complejo de aceleradores del CERN. Es un acelerador circular de 27 kilómetros de circunferencia, instalado en un túnel subterráneo, diseñado para poder acelerar las partículas hasta 7 TeV. Hasta el momento, se han alcanzado 6.5 TeV y la siguiente toma de datos se iniciará en 6.8 TeV

Dentro del acelerador, dos haces de hadrones (protones o iones pesados) de alta energía viajan cerca de la velocidad de la luz, en direcciones opuestas, dentro de dos tubos diferentes en vacío ultra alto (UHV). Estos haces se guían alrededor del anillo mediante un fuerte campo magnético generado por electroimanes superconductores, que se mantienen a una temperatura de $-271.3^{\circ}C$ mediante un sistema de distribución de helio líquido. En el LHC, los protones en los haces se hacen colisionar en paquetes o *bunches* de protones, dando lugar a gran cantidad de colisiones instantáneas.

En el Centro de Control del CERN se encuentran los servicios e infraestructura técnica del acelerador. Desde ahí, se hacen colisionar los haces en cuatro puntos del anillo acelerador, que corresponden con cuatro detectores de partículas: ATLAS, CMS, ALICE y LHCb. Estos son los cuatro experimentos principales del LHC, cuya diferencia fundamental está en sus detectores.

- ATLAS (A Toroidal LHC ApparatuS): es un detector de propósito general, cuyos objetivos abarcan desde la búsqueda del bosón de Higgs, hasta de nuevas dimensiones o partículas que conformen la materia oscura. Es, con el detector CMS, el experimento más grande del LHC.
- CMS (Compact Moun Solenoid): es también un detector de propósito general. Aunque comparte los objetivos científicos del ATLAS, emplea un diseño diferente.
- ALICE (A Large Ion Collider Experiment): dedicado a la física de iones pesados. Está diseñado para estudiar la física de la materia altamente interactiva a densidades de energía extremas, en las que se forma una fase de la materia llamada quark-gluon plasma.
- LHCb (Large Hadron Collider beauty): especializado en investigar las diferencias entre materia y antimateria estudiado un tipo especial de partícula: el quark *beauty* o quark *b*.

Además, el LHC cuenta con otros cuatro experimentos: TOTEM y LHCf, los más pequeños, se concentra en *partículas forward*, partículas generadas en direcciones prácticamente paralelas a la de los haces de protones o núcleos que colisionan. El experimento TOTEM cuenta con detectores en los extremos del punto de interacción de CMS, mientras que el LHCf los tiene a lo largo de la línea de los haces del LHC, a 140 metros a los lados del punto de colisión de ATLAS.

MoEDAL utiliza detectores cerca del LHCb con intención de encontrar una partícula hipotética, el monopolo magnético.

FASER, el experimento más reciente, se encuentra a 480 metros del punto de colisión de ATLAS y está dedicado a la búsqueda de nuevas partículas ligeras y al estudio de neutrinos [5].

El dispositivo experimental usado en este trabajo es el detector CMS, que se explicará con detalle a continuación, en la sección 2.3.

2.2. Magnitudes relevantes

A continuación, se presentan algunas magnitudes que aparecerán a lo largo de esta sección y en adelante, imprescindibles para la descripción del detector CMS.

El sistema de coordenadas empleado por el CMS toma como origen el punto de colisión de las partículas en el interior del experimento. El eje y apunta en la dirección vertical hacia arriba, y el eje z en la dirección del haz de partículas; el eje x es perpendicular a los dos anteriores, apunta de forma radial hacia el centro del LHC. El ángulo azimutal ϕ se mide desde el eje x en el plano x - y. El ángulo polar θ se mide desde el eje z. Así, se pueden definir las siguientes magnitudes, tomadas de [1]:

• Pseudorapidez η : indica la distancia de las partículas al plano transverso. Su expresión es:

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{2.1}$$

• Momento transverso p_T : momento de la partícula en el plano transverso:

$$p_T = \sqrt{p_x^2 + p_y^2} \tag{2.2}$$

Se emplea el momento en el plano transverso porque es donde se puede aplicar conservación del momento, pues al ser el plano perpendicular a la colisión el momento inicial era p = 0.

• Energía transversal faltante E_T^{miss} : se refiere a la energía que no es detectada tras una colisión, pero que, por las leyes de conservación de momento y energía, se esperaría encontrar. Esta energía se asocia a partículas, especialmente neutrinos, que no interaccionan por las fuerzas electromagnéticas o fuerte, y por ello, no se registran en el detector de partículas de un acelerador.

2.3. CMS

En esta subsección se va a describir en más detalle en detector usado para realizar los estudios presentados en este trabajo. Para ello, se ha empleado fundamentalmente el *CMS Physics: Technical Design Report, Volume I*, [1].



Figura 2.1: Detector CMS [1]

El detector CMS mide 22 m de largo y 15 m de diámetro, con un peso de 12.500 toneladas. En la figura 2.1 se puede ver un esquema del detector completo. En el interior del detector CMS se encuentra un solenoide superconductor de 13 m de longitud y 5.9 m de diámetro interno que produce campos magnéticos de hasta 4 T, magnitud necesaria para poder curvar la trayectoria de las partículas con carga eléctrica.

La curvatura de las partículas permite identificar su carga y medir su momento mediante la ley de Lorentz. Es decir, las partículas de carga opuesta se curvarán en direcciones opuestas y las de mayor momento se curvarán menos que las de momento bajo. En el centro del cilindro que conforma el solenoide se encuentra el punto de colisión de las partículas y a su alrededor se sitúan el resto de los componentes. El solenoide es lo suficientemente grande como para acomodar en su interior los calorímetros electromagnético y de hadrones y, dentro de estos, el *inner tracker* o detector de trazas. En el exterior del solenoide, rodeándolo, se encuentra el sistema de muones.

2.3.1. Sistema de detección de trazas

El inner tracker es la primera capa del detector. Se compone principalmente de sensores electrónicos de silicio de distintos tamaños. Este material es resistente a la radiación y, por sus propiedades de semiconductor, permite detectar partículas con carga eléctrica. Cuando una partícula cargada interacciona con el semiconductor, se genera una corriente eléctrica que se puede medir, indicando el paso de una partícula cargada por esa zona del detector. Su función es medir la trayectoria de las partículas que lo atraviesan y lo hace con una precisión de 10 μm . Se divide en tres partes repartidas a distintos radios con el objetivo de detectar las partículas de forma óptima en distintas condiciones de flujo:

- Pixel tracker: los detectores de píxel son los más cercanos al punto de interacción de las partículas ($r \approx 10 \ cm$), donde el flujo es mayor. El tamaño de un píxel es $\approx 100 \times 150 \ \mu m^2$.
- Zona intermedia: (20 < r < 55 cm), el flujo de partículas es más bajo, permitiendo usar detectores de silicio en forma de tiras de tamaño mínimo 10 cm × 80 μm .
- Zona externa: (r > 50 cm), el flujo se ha reducido lo suficiente como para usar tiras de mayor tamaño, con un máximo de $25 \text{ cm} \times 180 \mu m$.

Este diseño permite tener buena resolución espacial, que es clave para poder medir el momento de las partículas con gran precisión. Además, una buena resolución espacial es necesaria para identificar el vértice principal de la colisión, el punto donde la colisión inicial protón-protón tuvo lugar, y también posibles vértices secundarios generados por ejemplo en la desintegración de un hadrón B. El detector de trazas completo del CMS se muestra en la parte más interna de la figura 2.1, tiene un radio de aproximadamente 110 cm y una longitud de cerca de 540 cm.

2.3.2. Calorímetro electromagnético

El calorímetro electromagnético (ECAL) es una estructura hermética y homogénea, compuesta por 61200 cristales de tungstato de plomo, material centelleador, montados en el barril central y cerrados por otros 73224 cristales en los tapones o *endcaps*. La zona del barril central tiene un radio de 129 *cm* y cubre el intervalo de pseudorapidez de $0 < |\eta| < 1.479$, mientras que los tapones cubrirán $1.479 < |\eta| < 3.0$.

Los cristales de plomo están conectados a fotodiodos de avalancha de silicio (APDs) en el barril y a fototriodos de vacío (VPTs) en los tapones Así, cuando un electrón o fotón lo atraviesa, dejará una señal luminosa proporcional a la energía de la partícula que será registrada por el detector.

2.3.3. Calorímetro de hadrones

El objetivo del calorímetro de hadrones (HCAL) es detectar y medir la energía de los hadrones, partículas formadas por quarks y gluones que atraviesan el ECAL. Es fundamental para la detección de hadrones sin carga eléctrica, que no interactúan con el ECAL. El HCAL está construido a base de placas de latón que se intercalan con un material centellador. Se ha elegido el latón por su alta densidad, que hace que las partículas interactúen fuertemente con el material, generando una cascada de partículas secundarias de menor energía (cascada hadrónica) que llegan al material centellador y emiten luz. Esta señal detectable se transforma en una señal eléctrica que posteriormente se digitaliza. Con este método es posible medir tanto la posición de impacto como la dirección de la partícula, aunque la precisión aumentará con la energía de esta. El HCAL ha sido diseñado teniendo fundamentalmente en cuenta los parámetros del imán, ya que en su mayoría se encuentra dentro de la bobina, rodeando el sistema ECAL. Con intención de optimizar la contención y hermeticidad para la medida de E_T^{miss} , el diseño del HCAL maximiza la cantidad de material absorbente (latón, por ser absorbente y no magnético) dentro de la bobina, añadiendo también una nueva capa de centelleadores en el exterior, el Hadron Outer (HO).

2.3.4. Sistema de Muones

Los muones producidos en el centro del experimento son detectados en dos ocasiones. Al ser partículas cargadas, interactúan en primer lugar en el detector de trazas. A continuación, atraviesan el resto del detector a apenas sin interactuar, hasta que son frenadas en la parte más externa: las cámaras de muones. Las cámaras de muones son detectores gaseosos, por lo que el principio de detección se basa en la ionización del gas cuando pasa por éste una partícula cargada. Dado que las cámaras de muones se sitúan en la parte más externa del detector, se espera que sólo lleguen a ellas los muones y los neutrinos (que no son detectables por estas cámaras). Como los detectores anteriores, las cámaras de muones cuentan con una parte de barril, $|\eta| < 1.2$, y dos tapones $1.2 < |\eta| < 2.4$. Además, cada parte está compuesta por diferentes detectores, en función de la intensidad del campo magnético y del fondo de partículas en cada zona. El fondo de partículas se refiere a las partículas que no son muones, pero interactúan con el detector. En general, se habla de fondo de neutrones, o inducido por neutrones, ya que estos producen radiación gamma mediante interacciones nucleares, que dejan señales en las cámaras [7].

En la región del barril, donde tanto el fondo de neutrones como la tasa de muones son bajos, así como el campo magnético residual, se emplean cámaras **Drift Tube** (DT). Como se observa en la figura 2.2, el detector del barril consiste en 250 cámaras organizadas en cuatro estaciones (MB1, MB2, MB3 y MB4). Las cámaras de cada estación están escalonadas de forma que un muon de alto p_T que se genere cerca del borde de la sección cruce al menos 3 de las 4 estaciones. Cada estación está diseñada para devolver un vector del muon en el espacio, con una precisión de ϕ mayor que 100 μm en posición y que 1 mrad en dirección.

En la zona de los tapones, se tiene en cuenta que tanto la tasa de neutrones de fondo como la de muones son altas, además de un campo magnético alto, por ello se utilizan cámaras **Cathode Strip** (CSC). Entre los dos tapones se cuenta con 468 CSC organizadas en 4 discos perpendiculares al haz. Casi todas las CSC están colocadas de forma que exista solapamiento en ϕ , para evitar brechas en la detección de muones. La resolución espacial de cada una de cámaras es típicamente de 200 μm y la angular en ϕ del orden de 10 mrad.

Por último, ambas regiones se incluyen también cámaras **Resistive Plate** (RPC), también organizados en discos perpendiculares al haz de partículas. Estos detectores proporcionan una respuesta rápida, con buena resolución temporal, pero peor resolución de posición que las DT y CSC. A partir de los datos de las RPCs, se puede decidir si los datos se registran o no.



En total, el sistema de muones contiene aproximadamente 25000 m^2 de plano de detección activa, con casi un millón de canales electrónicos.

Figura 2.2: Esquema de un sector del sistema de muones.^[1]

Capítulo 3

Reconstrucción de objetos en CMS

Una vez registradas las interacciones de las partículas con el detector, ha de emplearse la información para reconstruir las partículas estables resultantes de una colisión. Para identificar las partículas tras la colisión se emplea un algoritmo iterativo que combina las medidas de cada capa del detector y recibe el nombre de **Particle Flow** (PF).

La figura 3.1 muestra la interacción de los distintos tipos de partículas con las diferentes capas del detector. En ella se ve que, como se expuso en el capítulo anterior, tanto los electrones como los fotones son frenados por el calorímetro electromagnético, los hadrones llegan hasta el calorímetro hadrónico y los muones son las únicas partículas detectadas que alcanzan las cámaras de muones. Además, se observa como el campo magnético permite determinar la carga de las partículas, según la curvatura de su trayectoria.

A continuación, se explican en detalle los criterios de selección del algoritmo Particle Flow.



Figura 3.1: Esquema de la interacción de las partículas con el detector CMS en el plano transverso. [8]

3.1. Particle Flow

La idea fundamental del algoritmo es combinar la información recogida por todos los subdetectores para lograr identificar todos los objetos resultantes de las colisiones. Asimismo, es importante obtener con la mayor precisión posible la medida de la energía y momento de cada partícula. Antes de combinar las señales de distintos subdetectores, el algoritmo trata de agrupar señales que puedan pertenecer a una misma partícula de forma local, es decir, en cada subdetector por separado, mediante algoritmos de reconstrucción de trazas y de agrupamiento. Posteriormente, une estos conjuntos de señales con los de otros subdetectores mediante un algoritmo de unión o *link*.

3.1.1. Reconstrucción local

Detector de trazas

En el detector de trazas se lleva a cabo una reconstrucción de forma iterativa. Primero se selecciona una traza con criterios muy restrictivos, con intención de minimizar la detección errónea. Una vez eliminadas las señales que se puedan asignar a esa traza sin ambigüedad, se repite el proceso, relajando progresivamente los criterios de selección. La relajación de estos criterios implica que la eficiencia en la detección será alta (no quedarán trazas sin reconstruir), pero al reducir la cantidad de datos en cada iteración se mantiene una tasa baja de detección errónea [6].

Calorímetros

En los calorímetros se emplea un algoritmo de agrupamiento independientemente en cada subdetector, ECAL y HCAL. Cada conjunto de señales agrupadas se denominará *clúster*. El primer paso del algoritmo es seleccionar la semilla del *clúster*, que será una celda del calorímetro cuya energía supera cierto umbral. A continuación, se consideran las celdas adyacentes y se añaden al *clúster* aquellas que tengan una energía lo suficientemente alta. Continuando de forma iterativa se construyen completamente los *clúster*.[6]

Sistema de Muones

El algoritmo de reconstrucción local en el sistema de muones consiste, como en el caso anterior en agrupar señales en DT, CSC o RPC individuales, formando *clústers* a partir de los que se pueda reconstruir un segmento de la traza de un muon. A continuación, se comienza con la estación más interna para la que haya un segmento reconstruido y se trata de buscar segmentos, o solamente señales, compatibles con este en otras estaciones de muones [11].

3.1.2. Algoritmo de unión (Link)

El objetivo de este proceso es reconstruir de forma completa la trayectoria cada una de las partículas, desde el origen de la colisión hasta que es completamente absorbida. De esta forma se evita además contar partículas por duplicado, en distintos subdetectores.

En general se espera que una partícula genere deposiciones en varios de los subdetectores. Por ejemplo, un hadrón cargado dejaría señales en el tracker y deposicones de energía en los calorímetros, mientras que un muon dejará señales en el traker y las cámaras de muones. Por lo tanto, el algoritmo se realiza sobre cada par de elementos con posibilidad de pertenecer a una partícula y se define una distancia para cuantificar la calidad de la unión. Así el algoritmo genera bloques, con uno, dos o tres elementos a considerar en la reconstrucción.

Para cada bloque, se procede como sigue: en primer lugar, se identifican los muones, ya que apenas interactúan con los calorímetros y su eliminación facilitará la tarea en estos subdetectores. A continuación, se reconstruyen los electrones, y por último los jets. La energía faltante una vez terminado el proceso, calculada por la conservación del momento, se asociará principalmente a la presencia de neutrinos u otras partículas invisibles.

Reconstrucción de muones

La identificación de muones según PF está diseñada para reconocer con la mayor eficiencia posible las siguientes partículas [16]:

- Muones *prompt*, aislados, que proceden de desintegraciones leptónicas de bosones W y Z.
- Muones de hadrones: tanto hadrones pesados, procedentes de desintegraciones de hadrones *beauty* o *charm*, como de hadrones ligeros, de desintegración de mesones $(q\bar{q}) \pi$ o K.

Además debe minimizar la identificación errónea de hadrones cargados, ya que estos se pueden reconstruir incorrectamente como muones si algún remanente de la cascada hadrónica alcanza el sistema de muones, lo que se conoce como perforación o *punch-through*.

La primera medida a considerar es la del detector de trazas. En él, se intentarán determinar el origen, el momento transverso y la dirección de las partículas cargadas. Para mantener una partícula en el análisis, han de encontrarse al menos dos señales en capas consecutivas del detector de píxeles y ocho en el total del detector. Como se ha visto anteriormente, la medida del detector de trazas será especialmente importante para muones de $p_T < 10 \ GeV$, para los que solamente se necesita una señal en las cámaras de muones.

Dependiendo de cómo se reconstruyan los muones atendiendo a sus deposiciones en los subdetectores, podemos clasificar los muones en tres categorías [16]:

Muones standalone: sólo se toma la señal en el sistema de muones. Las señales en los detectores DT y CSC se agrupan para formar segmentos de la trayectoria del muon. Este segmento se empleará como semilla en el reconocimiento de parámetros, en la búsqueda de todas las señales del muon en los DT, CSC y RPC.

La medida del momento de un muon *standalone* se determina por el ángulo de curvatura de este a la salida de la bobina, tomando como origen el punto de interacción.

• Muones del Tracker: si una señal en el detector de trazas tiene $p_T > 0.5 \ GeV$ y momento total $> 2.5 \ GeV$, se trata de extrapolar su trayectoria hasta el sistema de muones. Si se encuentra al menos una señal coincidente en las cámaras de muones, se identifica la partícula como muon del tracker.

Para muones de bajo momento, la mejor resolución será la obtenida del tracker de silicio.

• Muones globales: son muones standalone cuya trayectoria se puede extrapolar hasta una compatible en el detector de trazas. Es decir, su trayectoria se puede reconstruir completamente desde el detector de trazas hasta las cámaras de muones. Para muones de alto momento transverso, $p_T \gtrsim 200 \ GeV$, este modelo, que usa tanto la información del sistema de muones como del tracker, mejora la resolución del momento con respecto a la reconstrucción usando sólo la información del tracker.

Por otra parte, dependiendo del tipo de muon varían los criterios de selección adicionales basándose en sus propiedades. Para los muones globales, se comprueba si están aislados seleccionando señales adicionales en el tracker y depósitos de energía en el calorímetro. Se define una distancia angular $\Delta R < 0.3$, donde $\Delta R = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2}$. Si en un $\Delta R < 0.3$ no se encuentran otras partículas y la suma del p_T medido en el detector de trazas y la E_T de los depósitos no supera el 10 % del p_T del muon, éste se considera como aislado.

Para muones aislados no es necesario añadir ningún requerimiento adicional. Para muones que no superan los criterios de aislamiento debido a errores en la reconstrucción de los segmentos del *tracker*, puede ser identificado como muon *standalone* si se puede asociar a un número suficiente de señales en el sistema de muones (al menos 23 en DT o 15 en CSC, de 32 y 24, respectivamente). Por el contrario, si falla por ajustarse mal al modelo de muon global, se puede recuperar si se encuentran al menos 13 señales en el tracker.

Cuando un hadrón cargado es identificado por error como un mon, el algoritmo tiende a crear partículas neutras espurias para justificar los depósitos en el calorímetro. En caso de que un muon no sea identificado como tal, se tomará como hadrón cargado, incorporando como propia la energía de los depósitos cercanos de partículas neutras.

Los muones provenientes de jets, es decir, de desintegraciones hadrónicas, necesitan un criterio de selección más riguroso. Para estos muones no aislados, se requerirá la reconstrucción de un muón global con una χ^2 menor a 10, con al menos dos señales en el sistema de muones y cuya trayectoria del detector de trazas reconstruida a partir de al menos cinco capas, incluyendo una del detector de píxeles. Se añade la exigencia de que coincida al menos otro segmento de las cámaras de muones la señal del tracker, o que existan depósitos del calorímetro que sean compatibles con la traza reconstruida en el tracker.

De esta forma se eliminan la mayoría de los hadrones de alto p_T que podrían dar lugar a confusión, además de reducir las reconstrucciones erróneas de muones *standalone* a señales en el detector de trazas.

Se tomará como p_T del muon el medido en el detector de trazas, siempre que este sea menor a 200 GeV. Para valores mayores, se escogerá de acuerdo a la menor probabilidad χ^2 del los diferentes métodos de reconstrucción: reconstrucción usando solo el tracker, reconstrucción tomando información del tracker y el primer plano del detector de muones o reconstrucción global.

Reconstrucción de electrones

Los electrones son absorbidos por el ECAL, por lo que tan sólo se espera que su trayectoria contenga elementos del detector de trazas y *clústers* en el ECAL. Tras eliminar los muones, los *clústers* del ECAL corresponderán tan sólo a electrones, hadrones cargados o fotones. La identificación de electrones se realizará de forma similar a la de muones, cambiando las señales del sistema de muones por *clústers* en el ECAL y ajustando las exigencias en cada caso.

A los niveles de energía del LHC, la pérdida de energía de los electrones al interaccionar con la materia (*Bremsstrahlung*) es lo suficientemente alta como para alterar su trayectoria. Además, los fotones emitidos por este proceso también pueden ser detectados por el ECAL. Así, la identificación de electrones es un proceso con mayor probabilidad de error que el caso anterior.

Reconstrucción de jets

Por último, una vez extraídas del algoritmo las señales correspondientes a muones y electrones se tratarán de identificar los hadrones, diferenciando entre hadrones cargados y neutros. Comenzando a partir de una traza aún no identificada del *tracker*, se prolonga la trayectoria hasta las posiciones en las que se pueden esperar depósitos en el ECAL y HCAL. Si en las celdas esperadas los depósitos de energía fueran mayores a lo que se espera para la partícula (a partir de lo medido en el detector de trazas), se asumirá la existencia de fotones o hadrones neutros, según se encuentre en el ECAL y el HCAL. De la misma manera, si una vez identificadas todas las trazas del *tracker* existen aún *clústers* que no coinciden con ninguna partícula, se asignarán también a fotones, en el caso del ECAL, o a hadrones neutros, en el caso del HCAL [6].

Capítulo 4

Procesos Físicos de Estudio

Los muones no son partículas estables, sin embargo, su tiempo de vida ($2.1969811 \pm 0.000002 \cdot 10^{-6} s$ [19]) es lo suficientemente grande como para poder detectarlo cuando llega a las cámaras de muones. Por ser el LHC un colisionador hadrónico, se espera con más frecuencia hadrones en el estado final, sobre todo en colisiones elásticas, que no son de interés. Por ello, serán de especial interés aquellos procesos en los que se presenten leptones.

Los muones en el LHC se pueden generar a partir de procesos muy diversos, que dan lugar a muones con características muy diferentes. Por un lado, se tienen muones que provienen de la desintegración de bosones W o Z o del bosón de Higgs. Así, para estudiar dichos bosones y los procesos en los que aparezcan será de crucial importancia la detección de muones de este tipo. Por otro lado, se encuentran muones que provienen de la desintegración mesones: tanto de piones como de hadrones B.

Antes de comenzar, se define formalmente algunos conceptos importantes:

- La sección eficaz, σ , es la medida de la probabilidad de que un suceso ocurra tras una colisión (entre haces de partículas). Se expresa en unidades de área, en particular, el barn, 1 $b = 10^{-28}m^2$.
- La fracción de desintegración es la proporción de veces en que una

partícula inestable se desintegra dando lugar a otras partículas determinadas, siendo esto uno de sus posibles canales de desintegración. La suma de las fracciones de desintegración de una partícula suma uno.

4.1. Muones que provienen de bososnes

Los bosones $W, Z \neq H$ se desintegran mayoritariamente de forma hadrónica, no obstante, dadas las características del acelerador y el detector, los canales leptónicos son más limpios. En la figura 4.1 se muestran las fracciones de desintegración correspondientes a los bosones $W \neq Z$, mientras que la figura 4.2 muestra los del bosón de Higgs:



Figura 4.1: Fracciones de desintegración (en porcentaje) de los posibles canales de: (a) bosón W, (b) bosón Z [19].

$H ightarrow \gamma \gamma$	2.28×10^{-3}	$^{+5.0\%}_{-4.9\%}$
$H \rightarrow ZZ$	2.64×10^{-2}	$^{+4.3\%}_{-4.1\%}$
$H \to W^+ W^-$	2.15×10^{-1}	$^{+4.3\%}_{-4.2\%}$
$H \to \tau^+ \tau^-$	6.32×10^{-2}	$^{+5.7\%}_{-5.7\%}$
$H \rightarrow b \bar{b}$	5.77×10^{-1}	$^{+3.2\%}_{-3.3\%}$
$H \to Z \gamma$	1.54×10^{-3}	$^{+9.0\%}_{-8.9\%}$
$H \to \mu^+ \mu^-$	2.19×10^{-4}	$^{+6.0\%}_{-5.9\%}$

Figura 4.2: Fracciones de desintegración de los posibles canales del bosón H, con su incertidumbre relativa en la tercera columna [19].

Se observa en la figura 4.2 que el bosón de Higgs, se desintegra también en

bosones W y Z, que a su vez podrán desintegrarse en muones. La detección de muones de este tipo, en los canales $H \to ZZ$, $H \to W^+W^-$, y $H \to \tau^+\tau^-$, será de gran importancia en medidas del bosón de Higgs. Los muones *prompt* son también relevantes en la detección de quarks *top*, ya que este se desintegra mayoritariamente según $t \to W b$.

En esta sección se tratarán algunos de los procesos de desintegración leptónica de bosones que dan lugar a muones, en particular, los procesos Drell-Yan, la física de los quarks *top* y la física del bosón de Higgs.

4.1.1. Proceso Drell-Yan

Es un proceso conocido y bien caracterizados, con una sección eficaz notablemente alta. Por estas razones se utiliza para calibrar y comparar los datos con la simulación.

Se da en colisiones hadrón-hadrón, cuando tiene lugar un proceso de aniquilación entre un quark del primero y un antiquark del segundo. Entonces se crea un fotón o un bosón Z virtual, que se desintegran finalmente en un par leptón-antileptón. Drell-Yan se caracteriza especialmente por la producción de leptones de signo opuesto, se ejemplifica en 4.3. La energía de los quarks originales se transfiere casi por completo al par leptón-antileptón final.



Figura 4.3: Diagrama de Feymann de un proceso Drell-Yan

4.1.2. Procesos $t\bar{t}$

Siendo el quark top la partícula fundamental más pesada del modelo estándar, con una masa de 173.3 GeV/c^2 , es el único quark que se desintegra de forma semi débil, es decir, a un bosón W y un quark b en casi el 100 % de las ocasiones. Por el mismo motivo, su vida media es muy corta y se desintegra antes de poder hadronizarse [19].

Por todo ello, el quark top juega un papel fundamental en el estudio del modelo estándar y en áreas como la rotura de la simetría electrodébil (mecanismo de Higgs) o en la búsqueda de física más allá del ME.

En colisiones hadrónicas, los quarks top se producen de forma mayoritaria en pares top anti-top, $t\bar{t}$, a través de procesos $q\bar{q} \rightarrow t\bar{t}$ y $g\bar{g} \rightarrow t\bar{t}$. La producción $t\bar{t}$ es uno de los fondos principales en muchos procesos del ME o más allá del ME, por lo que es importante medir de forma muy precisa su sección eficaz.

De nuevo, de acuerdo con el modelo estándar, el quark top se desintegra a un bosón W y un quark b. De esta forma la producción de $t\bar{t}$, da lugar a estados finales con dos bosones W y dos jets procedentes de la fragmentación de los quarks b. Los estados finales de este proceso se pueden dividir en las tres categorías de la figura 4.4:

A.	$t\overline{t} \to W^+ b W^- \overline{b} \to q \overline{q}' b q'' \overline{q}''' \overline{b},$	(45.7%)
В.	$t\overline{t} \to W^+ b W^- \overline{b} \to q \overline{q}' b \ell^- \overline{\nu}_\ell \overline{b} + \ell^+ \nu_\ell b q'' \overline{q}''' \overline{b},$	(43.8%)

C. $t\overline{t} \to W^+ b W^- \overline{b} \to \ell^+ \nu_\ell b \ell'^- \overline{\nu}_{\ell'} \overline{b}.$ (10.5%)

Figura 4.4: Desintegración de pares $t\bar{t}$ [19].

Los quarks en los estados finales evolucionan a jets de hadrones. En caso de que ambos bosones W se desintegran de forma leptónica (C), el suceso contiene dos leptones con carga opuesta y gran momento, dos neutrinos (no detectados, pero medidos a partir de la energía transversal faltante) y al menos dos jets provenientes de un quark b. Este tipo de jets tienen una característica especial: vuelan antes de hadronizarse. Pueden identificarse

como provenientes de un quark b por presentar un vértice secundario. En la figura 4.4, se incluyen en paréntesis las contribuciones relativas asumiendo la universalidad leptónica. Así, debido a la poca contribución de otros procesos del ME, el canal dileptónico (C) es óptimo para medir la sección eficaz de los procesos $t\bar{t}$. En este caso, la ℓ hace referencia a cualquier leptón, pero generalmente el análisis separa entre los μ y e, y los τ , ya que este último es mucho más difícil de reconstruir. Al contrario que los e y los μ , los τ se desintegran dentro del detector tanto a otros e y μ , como a hadrones. Para las desintegraciones leptónicas, pasan a contar directamente como e o μ . En caso de desintegrarse a hadrones, existen también métodos para identificar su procedencia [3][19].

4.1.3. Producción y desintegración del bosón de Higgs

Por ser parte del mecanismo que otorga masa a las partículas, el acoplamiento del bosón de Higgs con las partículas fundamentales viene determinado por las masa de estas. Este nuevo tipo de interacción es muy leve para partículas ligeras como los quarks up y down o los electrones, e intensa para partículas como los bosones $W \ y \ Z \ y$ el quark top. De forma precisa, según el ME, la interacción del Higgs es linealmente proporcional a la masa de los fermiones y cuadráticamente proporcional a la de los bosones. Como resultado, los mecanismos dominantes para la producción y desintegración del Higgs involucran el acoplamiento de H a $W, \ Z \ y$ la tercera generación de quarks, (t, \bar{t}) y leptones, (τ^-, τ^+) [19].

Medidas de Higgs

Como se introdujo anteriormente, los modos de desintegración de H a bosones W Z y a leptones τ son esenciales para la medida del bosón de Higgs. Se presentan a continuación los canales de desintegración:

• $H \to ZZ^* \to l^+ l^- l'^+ l'^-$, donde $l, l' = e, \mu$. CMS emplea este canal, combinado con el $H \to \gamma\gamma$, especialmente para la medida de la masa del Higgs ya que tienen una alta resolución.

- $H \to W^+W^- \to l^+\nu l^-\bar{\nu}$, donde $l, l' = e, \mu$ y se pueden dar todas las combinaciones de sabores: $e^+e^-, \mu^+\mu^-$ y $e^\pm\mu^{\mp}$. Aunque la fracción desintegración correspondiente a este canal es más alta, la presencia de neutrinos resulta en una menor resolución en la masa de H.
- $H \to \tau^+ \tau^-$. En este canal se consideran τ que se desintegren a electrones, muones y hadrones. De nuevo, por la presencia de neutrinos, la resolución de la masa $m_{\tau^+ \tau^-}$ es baja.

Se puede ver de nuevo en la figura 4.2 que las fracciones de desintegración de estos canales son bastante menores que la del canal hadrónico $H \rightarrow b\bar{b}$. No obstante, los estados finales con leptones presentan una señal clara y proporcionan medidas con menos incertidumbre [19].

Existen diversos procesos que dan lugar a la producción de bosones de Higgs en colisionadores de altas energías. La producción por fusión de gluones (el de mayor sección eficaz), por fusión de bosones vectoriales o, entre otros, asociación con $t\bar{t}$. Este último tiene una sección eficaz baja, por lo que una selección precisa de muones será de gran importancia en la selección de sucesos de este tipo.

Producción de bosón de Higgs en asociación con $t\bar{t}$

Se ha visto que el acoplamiento del bosón de Higgs al resto de partículas fundamentales es proporcional a la masa de estas. Es entonces razonable esperar que el acoplamiento del bosón de Higgs con el quark top, la partícula fundamental más pesada, fuese mayor que el resto. El acoplamiento, o fuerza de interacción, del bosón de Higgs a otras partículas es generalmente proporcional a la probabilidad de que este se desintegre en un par de las mismas y así es como se mide para la mayor parte de partículas elementales. Sin embargo, el bosón de Higgs no puede desintegrarse en un par $t\bar{t}$, ya que estas partículas son demasiado pesadas (un solo quark top es casi 50 % más pesado que un H).

La forma alternativa de medir el acoplamiento del bosón es en su producción, un proceso también proporcional a la fuerza de interacción entre ambas partículas. Para ello, se buscan sucesos en los que tras la colisión se encuentren tanto un bosón de Higgs, como un par de quarks $t\bar{t}$. Hay principalmente dos situaciones que pueden dar lugar a este estado final, que se pueden ver en la figura 4.5:

- a. Se forman dos pares de quarks top independientes. Un quark y un antiquark provenientes de distintos pares interaccionan, dando lugar a un bosón de Higgs. El resultado es un bosón H y un par $t\bar{t}$
- b. Se genera tan sólo un par $t\bar{t}$, y el bosón de Higgs es radiado por uno de los quarks.

Este proceso recibe el nombre de *Producción del bosón de Higgs en asocia*ción con un par top-antitop, o $t\bar{t}H$, y constituye aproximadamente tan sólo el 1% de la producción de bosones de Higgs, un fenómeno ya poco frecuente de por sí [19].



Figura 4.5: Diagramas de Feymann para la producción del bosón de Higgs en asociación con un par $t\bar{t}$.

4.2. Muones que provienen de mesones

Los hadrones formados por pares quark- antiquark se conocen como mesones. Los muones que provienen de la desintegración de mesones, en general, tienen una energía más baja que los muones provenientes de bosones. Asimismo, puesto que los mesones vuelan antes de desintegrarse, los muones se generarán a una cierta distancia del vértice primario y no en el centro del detector. Los procesos principales en esta categoría serán la desintegración a muones de piones o de hadrones B. Se exponen algunos casos de interés.

4.2.1. Desintegración de hadrones B

Los hadrones que contienen un quark *bottom*, b, se conocen como hadrones de sabor b o hadrones B. En este caso, nos interesaremos por ciertas desintegraciones de hadrones B a leptones, o más en particular, a muones.

Desintegraciones semileptónicas.

Las desintegraciones semileptónicas de los B-hadrones proporcionan información importante a la hora de medir con precisión los valores de elementos $|V_{ub}| \ge |V_{cb}|$ de la matriz CKM, que contiene información sobre la intensidad de las desintegraciones débiles en las que se produce cambio de sabor. Dos vías para determinar estos elementos de matriz son las transiciones:

$$b \to c \ l \ \bar{\nu}_l \quad y \quad b \to u \ l \ \bar{\nu}_l \quad con \quad l = e, \mu.$$

Para ello se emplearán la medida de la tasa de desintegraciones, $\bar{B} \to X \ l \ n\bar{u}_l$, sumando todos los posibles estados hadrónicos, o la tasa de desintegraciones exclusivas, donde el estado hadrónico final es un mesón concreto $(X = D, D^*, \pi, etc.)$ [19].

Desintegración de baja probabilidad.

Los procesos de desintegración de b-hadrones de baja probabilidad en el ME sólo pueden darse a través de corrientes neutras de cambio de sabor (*Flavour-changing neutral current*, FCNC). Además, es necesario que se produzca una aniquilación interna de los quarks de los B-mesones. Por ello, es posible que los efectos de nueva física tengan una gran contribución en estos
procesos y su estudio es fundamental. Las transiciones principales en esta categoría son: \mathbb{P} = \mathbb{P}^0 = \mathbb{P}^0

$$B_s \to \mu^+ \mu^- \quad y \quad B^0 \to \mu^+ \mu^-$$

Hasta el momento, según el ME, se conoce que la fracción de desintegración de estos procesos es el siguiente, pero existen nuevas teorías que estiman una fracción de desintegración superior, por lo que la selección de muones de este tipo será fundamental para continuar con su estudio.

$$\mathcal{B}^{ME}(B_s \to \mu^+ \mu^-) = (3.66 \pm 0.23) \times 10^{-9}$$
$$\mathcal{B}^{ME}(B^0 \to \mu^+ \mu^-) = (1.06 \pm 0.09) \times 10^{-9}$$

Capítulo 5

Introducción al Machine Learning

El Machine Learning (ML) o aprendizaje automático es una disciplina que emplea algoritmos para conseguir que los ordenadores identifiquen patrones en un conjunto de datos y que realicen predicciones, a partir de estos, en otros conjuntos. El conjunto de datos a partir de los que el ordenador aprende se denomina *set de entrenamiento*. Un modelo de ML aplicado a un gran set de entrenamiento o a problemas complejos, por ejemplo, de varias dimensiones, puede detectar patrones que no eran si quiera apreciables por el programador. A este proceso se le llama *data mining*.

A no ser que se indique lo contrario, el desarrollo de este capítulo se basará en el libro Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems, de Aurélien Géron, [9].

5.1. Sistemas de Machine Learning

Existen diferentes tipos de sistemas de Machine Learning y se pueden clasificar según los siguientes criterios:

- Si se entrenan o no con supervisión humana.
- Si pueden o no incrementar el aprendizaje gradualmente, sobre la marcha: aprendizaje online o tipo *batch* ("en lote").
- Si trabajan comparando nuevos datos con datos conocidos o si detectan patrones en los datos de entrenamiento y construyen un modelo predictivo: *instance-based learning o model-based learning*.

5.1.1. Aprendizaje supervisado vs no supervisado

La clasificación según la cantidad y tipo de supervisión que tiene lugar durante el entrenamiento, se diferencian cuatro categorías: aprendizaje supervisado, no supervisado, semi supervisado o reforzado.

Aprendizaje supervisado

Se considera aprendizaje supervisado cuando el set de entrenamiento que se proporciona al algoritmo incluye las soluciones deseadas, *etiquetas*. Una de las tareas principales de aprendizaje supervisado es la clasificación. Un ejemplo claro es el filtro de spam, que tras entrenarse con gran cantidad de emails de clase ya determinada (spam o no spam), debe aprender a clasificar los nuevos.

Otra tarea típica es la de predecir un valor numérico "objetivo" a partir de un conjunto de características dadas, *predictores*. Esta tarea recibe el nombre de *regresion*. Un ejemplo sería el precio de un coche, a partir de datos como la antigüedad, marca o kilómetros recorridos. Los modelos utilizados para tareas de regresión pueden también usarse para clasificación y viceversa.

Aprendizaje no supervisado

En contraposición a los modelos anteriores, en este caso los datos de entrenamiento están sin etiquetar.

Algunas de las tareas más relevantes a las que se aplican este tipo de modelos son el *clustering* (o agrupamiento) y la detección de anomalías. Los modelos de *clustering* dividen los datos en número de grupos tal que los datos de cada grupo sean más similares entre ellos que a los de otros grupos. En la detección de anomalías el algoritmo se entrena únicamente con ejemplos considerados normales y, al ver una nueva puede determinar si es normal o no lo es.

Aprendizaje semisupervisado

Estos algoritmos se emplearán en casos en los que haya una mayoría de datos no etiquetados, pero también algunos ejemplos etiquetados. En general, son una mezcla de algoritmos supervisados y no supervisados. Algunos servicios de almacenamiento de imágenes hacen uso de modelos de este tipo: detectan que una persona aparece en un conjunto de imágenes y las agrupa (parte no supervisada, *clustering*), entonces requiere que el usuario etiquete tan sólo una de las imágenes con el nombre de la persona, y con ello etiquetará todas las imágenes del grupo.

Aprendizaje reforzado

Este es el modelo que más difiere de los anteriores. En este caso el sistema que "aprende", denominado *agente*, debe observar el entorno y decidir qué acciones realizar, tras lo que recibe refuerzo positivo o negativo. Así, desarrolla la mejor estrategia posible, *táctica*, para maximizar el refuerzo positivo y minimizar el negativo. Se implementan, por ejemplo, en robots que deben aprender a caminar.

5.1.2. Aprendizaje online vs tipo batch

Otro modo de clasificar los algoritmos de ML es según si pueden o no incrementar el aprendizaje sobre la marcha.

Si el sistema es incapaz de incrementar el aprendizaje, *Batch Learning*, este debe entrenarse empleando todo el set de datos disponible y una vez entrenado y lanzado a producción, aplicará lo que ha aprendido sin aprender nada más. En caso de necesitar que el sistema aprenda de nuevos datos, habría que entrenarlo desde cero en el set de datos completo, los antiguos y los nuevos.

Por el contrario, en el aprendizaje *Online*, el sistema se entrena pasándole datos de forma secuencial. Cada paso de aprendizaje es más rápido y menos costoso computacionalmente, además de que se ahorra espacio de almacenamiento. Este método es más útil para sistemas que reciben nuevos datos continuamente, como la bolsa. También son útiles para entrenar algoritmos en bases de datos tan grandes que no caben en la memoria de una sola máquina, *out-of-core learning*.

5.1.3. Aprendizaje basado en instancias vs basado en modelos

Esta clasificación está basada en la forma en la que el algoritmo generaliza lo aprendido a partir del set de entrenamiento a los nuevos casos.

El aprendizaje basado en instancias consiste en que el algoritmo aprende los ejemplos de "memoria" y utilizando una *medida de similitud*, los compara con los nuevos casos. El aprendizaje basado en modelos trata de hacer predicciones, usando los ejemplos del set de entrenamiento para construir un modelo predictivo. Se trata de identificar una tendencia en los datos y seleccionar un modelo que parezca describirla y a continuación deben ajustarse los parámetros de este. Para ello se pueden definir funciones de rendimiento del modelo, que compruebe como de bueno es, o una función de coste que mida como de malo es.

5.2. Modelos usados en este trabajo

La tarea que se llevará a cabo en este trabajo es de clasificación, por lo que todos los modelos que se han usado serán de aprendizaje supervisado, en *batch* y basados en modelos.

Las partes principales de los modelos supervisados serán:

- Variable objetivo: Y, lo que queremos predecir.
- Variables de predicción: $\{X_1, X_2, ..., X_n\}$, las variables de entrada que se emplearán para hacer la predicción.
- Modelo de predicción: $Y = F(X_1, X_2, ..., X_n)$ el "motor de predicción" que estimará el valor de la función.

En los modelos supervisados de clasificación el conjunto de datos etiquetados se divide en dos, el set de entrenamiento y el set de comprobación (*train-test*). En general, se emplearán un 70 % - 80 % de los datos para entrenar el modelo y el 30 % - 20 % restante para ponerlo a prueba.

El problema fundamental del entrenamiento de los algoritmos de clasificación es encontrar un modelo que aprenda lo suficiente de los datos de entrenamiento como para poder hacer predicciones correctas, pero que no se ajuste demasiado a ellos, aprendiendo de fluctuaciones estadísticas, dando lugar a un modelo muy complejo que no prediga adecuadamente en otros sets de datos, lo que se conoce como sobre entrenamiento u *overfitting*. Se ejemplifica en la figura 5.1.



Figura 5.1: Ejemplificación de un modelo de clasificación con falta de entrenamiento, apropiado, y con sobre entrenamiento.

Existen diversos métodos para comprobar la eficacia de un modelo, aunque en este trabajo se emplearán curvas ROC, (*receiver operating characteristic curve*, curva característica de funcionamiento del receptor), y matrices de confusión.

 Curvas ROC: es una gráfica que ilustra la capacidad de diagnóstico de un modelo de clasificación binario. Se crea representando la tasa de verdaderos positivos (TVP) frente a la tasa de falsos positivos (TFP). La TVP se define como la frecuencia en la que el modelo predice un resultado es positivo (señal) cuando la etiqueta es positiva (señal):

$$TVP = \frac{VP}{VP + FN}$$

Donde VP representa el número de verdaderos positivos y FN el de falsos negativos, es decir, cuando la etiqueta es positiva pero la predicción negativa. La TVP se conoce también como la eficiencia del modelo, y en adelante se empleará este término.

La TFP se define como la frecuencia en la que el modelo predice un resultado es positivo (señal) cuando la etiqueta es negativa (fondo):

$$TFP = \frac{FP}{FP + VN}$$

Donde, análogamente, FP representa el número de falsos positivos (etiqueta negativa y predicción positiva, fondo como señal) y VN el de verdaderos negativos (etiqueta negativa y predicción negativa, fondo como fondo).



Figura 5.2: Ejemplo curva ROC

Un modelo eficaz debería devolver una eficiencia alta y una TFP baja, es decir, el área bajo la curva ROC debería aproximarse a 1.

 Matriz de confusión: es una tabla que se emplea para mostrar el rendimiento de un modelo de clasificación en el set de comprobación. Se representan mediante una tabla de doble entrada el número de elementos clasificados en cada categoría según las etiquetas frente a lo predicho por el modelo. Es decir, representa las tasas VP, VN, FP, FN como se ve en la imagen 5.3.



Figura 5.3: Ejemplo matriz de confusión

A partir de estos valores se definen:

- La **precisión** del modelo: (VP + VN)/Total
- El **error** del modelo: (FP + FN)/Total

5.2.1. Árboles de decisión

Los árboles de decisión son el sistema más simple de los que se emplearán para la clasificación. Se trata de clasificar la variable objetivo Y basándose en un conjunto de variables de predicción $\{X_1, X_2, ..., X_n\}$.

Su estructura, expuesta en la figura 5.4, es la siguiente: cada **nodo** corresponde a una "prueba" sobre un atributo, y cada **rama** que sale de él será un posible valor del atributo. Los nodos finales, **hojas**, representan una clase final, y el **camino** que lleva a ellas será el conjunto de los valores de los atributos en esa clase.

Los árboles de decisión se pueden construir relativamente rápido, dando resultados generalmente de eficacia similar a otros más complejos. Además, por su representación tan intuitiva, se interpretan fácilmente por los humanos.



Figura 5.4: Ejemplo de árbol de decisión para decidir si salir a hacer deporte en el exterior. En rojo se marcan los nodos de decisión, en verde las clases finales y en azul los posibles valores de los atributos medidos en los nodos.

Los algoritmos para construir árboles de decisión difieren principalmente en el criterio que utilizan para realizar la división en los nodos, *splitting*, aunque la idea de todos ellos es evaluar cada atributo según su poder de separación, e incluir los más potentes en la decisión. Es decir, si un atributo no es útil para la clasificación, no se usará para aumentar el árbol. Esta característica hace que los árboles de decisión sean útiles como preprocesadores de otros modelos de ML a los que perjudique la presencia de información

innecesaria.

Al construir el modelo se han de fijar distintos **hiperparámetros**, como el criterio de selección, la profundidad del árbol, el número de hojas, el número de elementos por hoja, etc. Estos valores determinarán el rendimiento del árbol y, aunque un modelo simple puede no ser perfectamente eficaz en la clasificación, un modelo complejo (por ejemplo, un árbol con demasiada profundidad), dará lugar a sobre entrenamiento. El objetivo es ajustar los hiperparámetros para obtener la mayor precisión posible, sin sobre entrenamiento.

A partir de los árboles de decisión se construyen los métodos de aprendizaje conjunto o *ensamble learning*, cuya idea fundamental es combinar múltiples modelos débiles para crear un modelo que mejore a los individuales tanto en estabilidad como en rendimiento.

Existen dos métodos para llevarlo a cabo: **bagging** y **boosting**.

5.2.2. Random Forest

Un algoritmo de machine learning que emplea el método bagging es el Random Forest (RF).

El primer paso para construir un modelo de aprendizaje conjunto es dividir el set de entrenamiento original en distintos subconjuntos, para entrenar un árbol de decisión en cada uno. Para ello se emplea el **bootstrapping**, es decir, se generarán M subconjuntos del set de entrenamiento tomando aleatoriamente elementos del conjunto original, con reemplazo. Entonces, se entrenarán M árboles de decisión en paralelo y, aunque estos tendrán un sesgo muy bajo, habrá una gran variabilidad entre ellos. La predicción final en el set de prueba se dará teniendo en cuenta la decisión de cada árbol, por ejemplo, calculando la media o por voto mayoritario.

Debido a que en los árboles de decisión se examinan todas las variables de predicción (p) al hacer una división, los M árboles distintos tendrán una

estructura similar, devolviendo predicciones correlacionadas. Para generar submodelos que no estén correlacionados, en los Random Forest se limita el número de variables a considerar en cada división, escogiéndose m variables de forma aleatoria. Un valor típico será $m = \sqrt{p}$.

Cuando se construyen los árboles a partir de los subsets "bagged", se puede calcular cuánto se reduce el error para una variable en cada punto de división. Tomando la media de esta reducción entre todos los árboles, se puede estimar la **importancia** que cada variable de input tendrá en la clasificación.



Figura 5.5: Esquematización del funcionamiento de un modelo de ML tipo Bagging [13].

5.2.3. Boosted Decision Trees

En contraposición a los modelos bagging, en los que los árboles se entrenan en paralelo, existen los modelos **boosting** en los que el proceso se realiza de forma secuencial. Es decir, en los modelos boosting, los resultados del primer árbol entrenado decidirán en que se debe concentrar el modelo siguiente, y así sucesivamente hasta que se hayan entrenado los M modelos. En adelante, se hará referencia a los modelos de Boosted Decision Trees como BDT.

En este caso, los datos del set de entrenamiento tendrán un peso asociado que se actualizará tras el entrenamiento de cada árbol, por lo que al realizar el bootstrapping no todos los elementos tendrán la misma probabilidad de ser seleccionados. La forma principal de enfocar el árbol siguiente hacia una mejor precisión es atribuyendo un peso mayor a los elementos que hayan sido clasificados erróneamente.

Además, el algoritmo asociará un peso a cada árbol resultante, en función de su rendimiento individual, y así la clasificación de los datos nuevos se basará también en todos los submodelos, pero dándole más importancia a los que obtuvieran mejor resultado. En algunos modelos de este tipo se añade una condición extra para decidir si mantener o rechazar cada árbol, como por ejemplo que tengan un error menor al 50 %.

Al asociar un peso mayor a los árboles que funcionen mejor en el set de entrenamiento, los métodos de boosting tienen mayor tendencia al sobre entrenamiento que los de bagging, aunque mayor capacidad de reducir los sesgos.



Figura 5.6: Esquematización del funcionamiento de un modelo de ML tipo Boosting [13].

Capítulo 6

Análisis multivariante para la clasificación muones

El análisis multivariante, o MVA, se define como la rama del análisis estadístico que se centra en la investigación simultánea de dos o más características (observables físicos) medidas en un conjunto de objetos [10].

En el contexto de este trabajo, el análisis multivariante consistirá en la aplicación de las técnicas de ML expuestas a la clasificación de muones. A lo largo de este capítulo se comentarán la muestra empleada para llevar a cabo el análisis, las variables disponibles y las variables de salida del MVA.

6.1. Simulaciones de Monte Carlo

El desarrollo, mantenimiento y perfeccionamiento del acelerador de partículas LHC, y el de sus detectores, se realiza empleando datos obtenidos mediante simulaciones de Monte Carlo. Las simulaciones de Monte Carlo tienen como objetivo recrear de la forma más exacta posible todos los procesos físicos que tienen lugar durante las colisiones. Es decir, procuran reproducir el experimento en cuestión, para perfeccionar su estudio. Para la generación de los sucesos simulados, CMS emplea un proceso que consta de tres programas:

El generador de sucesos de Monte Carlo simula colisiones protón-protón o entre iones pesados, con una cierta energía de centro de masas, la que se quiera según las condiciones del LHC. En el caso de este trabajo se usarán colisiones protón-protón y una energía en centro de masas de 13 TeV. El objetivo de este *software* será simular las características físicas de la partícula antes de su interacción con el detector. Dentro de la generación de sucesos, hay que simular dos procesos diferentes:

- El hard-scattering, que es la colisión principal: se muestran las características físicas de los protones, se elige el momento de los partones (componentes de los hadrones) que van a colisionar y se simula según cálculos de teoría de perturbaciones el proceso deseado. Entre los tipos de software que se emplean para esto destacan aMC@NLO o Powheg.
- Habrá una emisión radiativa debida a las partículas que radian gluones y fotones, esto se conoce como radiación de estado inicial y final y se modela con *PYTHIA*.

A continuación, se reproducen la propagación de las partículas y su interacción con el detector. En este punto de la simulación, se pueden emplear dos aproximaciones distintas:

- Simulación completa: el paquete OSCAR (Object oriented Software for CMS Analysis and Reconstruction) computa la propagación de las partículas generadas en el detector e incluye las interacciones de estas con todas sus componentes. OSCAR está basado en Geant4 y contiene descripciones precisas de la geometría del detector y sus materiales. Esta aproximación requiere de un alto coste computacional.
- Simulación rápida: el paquete FAMOS (FAst MOnte- Carlo Simulation), programado en C++, consiste en una simplificación del modelo anterior. Se incluyen en la simulación las interacciones de mayor importancia de las partículas con la materia. Con FAMOS, toda la cadena de simulación, desde la generación de sucesos hasta el análisis, se realiza en un solo paso.

Por último, la simulación de las respuestas electrónicas de salida, diagonalización, se lleva a cabo por el programa de reconstrucción ORCA (Object oriented Reconstruction for CMS Analysis).

En este proceso, se imponen sucesos de pile-up a las colisiones protónprotón de alta energía. Los sucesos de pile-up hacen referencia a las colisiones protón-protón secundarias que tienen lugar cuando se cruzan dos paquetes de protones o *bunches* que se superponen a la interacción principal y complican la reconstrucción de objetos [1][2].

6.1.1. Muestra empleada en este trabajo

Para el entrenamiento de todos los modelos que se expondrán, se ha empleado una muestra de $t\bar{t}$ que reproduce los procesos explicados en la sección 4.1.2. Cabe recordar que en los procesos $t\bar{t}$ la desintegración será principalmente a bosones W y quarks b, por lo que la muestra simulará un escenario bastante realista, con gran variedad de muones.

Para el entrenamiento será imprescindible decidir cuáles son los sucesos que se quieren clasificar. En este trabajo se tomarán como señal (muones a clasificar o "buenos") los muones que provengan de bosones, τ , de hadrones B y otros hadrones ligeros.

Como fondo se tomarán los sucesos reconstruidos como muones que a nivel de generación no se pueden asociar a ninguna partícula (not matched) y los hadrones que hayan dejado señales en las cámaras de muones (*punchthrogh*).

Para observar las diferencias entre estos tipos de sucesos, se genera la variable *matching* a partir de la información de la simulación de Monte Carlo. Del primer paso de la simulación, la generación, se conoce la fuente de la partícula tras la colisión. A continuación, se obtiene la información de su paso por cada uno de los subdetectores, que se puede reconstruir mediante los algoritmos habituales. La variable *matching* se crea uniendo la información de generación de la partícula con la de la reconstrucción, tras simular su paso por el detector. Para el análisis se ha tomado como fondo todo suceso con *matching* ≤ 2 , y como señal *matching* > 2. Dentro de la definición de señal

cabe mencionar que los muones provenientes de bosones se corresponden con matching = 10. En la figura 6.1 se ve cómo se distribuyen el número de sucesos de cada tipo y como, en la definición de señal, predominan los muones de bosones.



Figura 6.1: Variable *matching* de los sucesos en la muestra $t\bar{t}$

La figura 6.1 muestra que hay claramente una cantidad considerablemente mayor de sucesos de señal que de fondo. No obstante, hay suficientes sucesos de cada tipo para llevar a cabo el estudio. En la tabla 6.1, se muestra el número de sucesos de cada tipo en el set de entrenamiento y en el de test.

	Entrenamiento	Test
Muones de bosones	17933675	23911126
Muones de mesones	18422341	24563563
Fondo	2103200	1402134
Total	38459216	49876823

Tabla 6.1: Número de sucesos de cada tipo en el set de entrenamiento y en el de test.

6.2. Variables de entrada del MVA

Una vez determinada la definición de señal y fondo, es necesario escoger las variables que se incluirán en el entrenamiento, es decir, las que los modelos empleará para clasificar los sucesos. Durante el proceso de reconstrucción, se puede obtener gran cantidad de información para cada objeto a reconstruir. El objetivo de esta sección será encontrar aquellas que puedan suponer un buen discriminante, las que presenten mayores diferencias entre señal y fondo.

Las variables disponibles para el MVA serán, [15]:

- *Global muon*: variable booleana que compruebas si el muon ha pasado los criterios de selección de muon global.
- χ^2 normalizada: compatibilidad de la reconstrucción global si el muon es identificado como muon global, o del detector de trazas si el muon se identifica como sólo del tracker (tracker only).

En las variables que representen una χ^2 , un valor alto de la variable implicará una baja compatibilidad.

- χ^2 *local*: compatibilidad de la reconstrucción del tracker y de la reconstrucción standalone.
- *Valid hits*: número de señales en las cámaras de muones incluidas en el ajuste de muon global, si el muón se reconstruye como global, o en el ajuste de muon del tracker, si se reconstruye como tracker only.
- Kink: χ^2 del kink-finder en el detector de trazas. El kink-finder es un algoritmo que parte la trayectoria del muon reconstruida en el tracker y la divide en dos en distintos puntos. En cada punto calcula la compatibilidad de las trazas.
- *Matched stations*: número de estaciones de las cámaras de muones con trazas.
- Segment compatibility: compatibilidad entre los segmentos de las cámaras de muones.
- Valid Fraction: fracción de señales válidas en el tracker.

- *Valid Pixel*: número de señales válidas en el píxel tracker (capa más interna del detector de trazas).
- Tracker Layers: número de capas del detector de trazas con señales.
- p_T : momento transverso de la partícula.
- η : pseudorapidez de la partícula.
- *nVtx*: número de vértices. Además del vértice principal, que genera la colisión del suceso a estudiar, el número de vértices que se pertenecen a la misma colisión entre paquetes de protones.
- *dxy*: parámetro de impacto en el plano X-Y.
- *dz*: parámetro de impacto en la dirección z.

En las figuras siguientes, 6.2, 6.3 y 6.4, se muestran las distribuciones de cada una de las variables presentadas para los sucesos de señal y los de fondo. Es lógico esperar que aquellas variables que presenten mayores diferencias entre ambas distribuciones sean de mayor importancia para la clasificación.

Algunas variables como la compatibilidad de los segmentos, 6.2e, la χ^2 normalizada 6.3a, o el número de estaciones con señales 6.2d, entre otras, presentan diferencias considerables, por lo que serán fundamentales en el entrenamiento.



Figura 6.2: Variables de entrada en el MVA: (a) *Global Muon*, (b) *Number of Valid Hits*, (c) *Tracker Layers*, (d) *Matched stations*, (e) *Segment Compatibility*, (f) *Valid Fraction*.



Figura 6.3: Variables de entrada en el MVA: (a) χ^2 normalizada, (b) χ^2 local, (c) Kink, (d) p_T , (e) dz, (f) dxy.



Figura 6.4: Variables de entrada en el MVA: (a) Valid Pixel, (b) η

Para el estudio completo de un MVA, será importante tener en cuenta no sólo la eficiencia del modelo, si no también los parámetros y conceptos físicos que estamos estudiando, de forma que este se adapte a ellos de la mejor forma posible. Por ello, algunas de las variables han de estudiarse en mayor detalle. En particular, se estudiará p_T , 6.3d, y η , 6.4b en la sección 6.2.1 y dxy, 6.3f, dz, 6.3e en la sección 6.2.2.

6.2.1. Momento transverso p_t y pseudorapidez η

Se ha explicado en el capítulo 4 de procesos físicos de estudio que por norma general los muones que provengan de bosones, muones *prompt*, llegarán a tener un valor de p_T mucho más alto que los que provengan de mesones. Es poco habitual encontrar muones *prompt* con valores bajos de p_T , mientras que los de mesones se encontrarán sobretodo en esta región.

Asímismo, como se muestra en la figura 6.5a, la distribución en p_T de los muones de mesones es muy similar a la del fondo, por lo que, si se incluyese esta variable directamente en el MVA, es muy probable que el modelo no los clasificase como muones.

Claramente, se tienen dos distribuciones muy distintas de p_T en la muestra

de muones y, como objetivo principal, nos interesa que el modelo clasifique ambas como señal.

Una manera de solventar esta situación es introducir la variable "repesada" en en MVA. Se asigna un nuevo peso al p_T de cada suceso, de forma que se mantenga la forma de su distribución y el modelo pueda aprender de la correlación que existe entre esta y el resto de variables, pero que no se decida en base a ella la clasificación.

En el caso de la variable η , 6.5b, se ve que no existe esta diferencia drástica entre los muones de bosones y de mesones. Sin embargo, su diferencia con la distribución del fondo se debe a un efecto de la reconstrucción y no es de interés que el modelo aprenda de ella.

Todos los modelos se entrenarán con p_T y η repesados. Además, se ha entrenado un modelo que no las incluye como variables de entrada, que servirá para comprobar si el repesado está condicionando el aprendizaje tal como se pretende.



Figura 6.5: p_T y η para las distintas contribuciones a la muestra: las señal, con muones de bosones y muones de mesones, y el fondo. (a) p_T , (b) η .

Además del repesado de estas variables, la variable p_T tendrá relevancia en otro aspecto del estudio. La distinción del origen de los muones es bastante significativa en p_T , tal como se ha comentado y se muestra en 6.5a. Por lo tanto, una forma de optimizar el entrenamiento del modelo para la clasificación de muones que provengan únicamente de bosones será restringiendo la muestra a aquellos sucesos que tengan un $p_T > 20 \text{ GeV}$. Al imponer esta selección estaremos asegurando que el modelo no aprenderá de los muones de bajo p_T , la mayoría provenientes de mesones, por lo que se optimiza la selección para muones de bosones.

En el capítulo 7, mostraremos cómo afecta esto al comportamineto del modelo, frente al de los modelos entrenados para $p_T > 10 \ GeV$, más inclusivos. Para valores menores de p_T no es probable encontrar muones provenientes de bosones, por lo que, en general, se aplicará la selección $p_T > 10 \ GeV$ en la muestra original.

6.2.2. Parámetros de impacto, dxy y dz

Las figuras 6.3f y 6.3e, que representan respectivamente dxy y dz, parecen presentar diferencias claras en las distribuciones de señal y fondo. No obstante, si tenemos en cuenta las distintas contribuciones de señal en la muestra, se espera que exista diferencia entre los parámetros de impacto de muones *prompt* y muones de mesones.

Como se expuso en el capítulo 4, los muones originados por desintegraciones de hadrones se caracterizan porque se generan a cierta distancia del centro del detector. Es decir, los hadrones de los que proceden se desplazarán del vértice primario antes de desintegrarse, dando lugar a mayores valores de dxy y dz que los muones *prompt*. La figura 6.6 lo ejemplifica claramente.

Por esto, la decisión de incluir o no las variables dxy y dz debe tomarse en función del tipo de muones para los que se pretende optimizar la clasificación.

No incluir los parámetros de impacto implicará que no se genere la distinción entre tipos de muones basada en estos observables, ambas contribuciones tendrán alta probabilidad de clasificarse como señal. Si se introducen, por el contrario, el MVA podrá aprender mayores diferencias entre los dos tipos de muones, siendo más propenso a clasificar los muones de mesones como fondo.



En en capítulo 7, se exponen modelos ejemplificando cada caso.

Figura 6.6: Parámetros de impacto para las distintas contribuciones a la muestra: las señal, con muones de bosones y muones de mesones, y el fondo. (a) dxy, (b) dz

6.3. Variable de salida del MVA

Es imprescindible definir, una vez entrenados los modelos, qué valores se utilizarán para estudiar su comportamiento de forma que se puedan comparar coherentemente.

Dos figuras de gran importancia para observar el funcionamiento general de la clasificación serán las expuestas en la sección 5.2: la matriz de confusión, que muestra el número de sucesos en cada categoría, y la curva ROC, que representa la eficiencia del modelo frente a la tasa de falsos positivos (TFP).

La distribución de salida del MVA, el discriminante, toma valores entre 0 y 1, y cuanto más alto es el valor para un suceso, más probable es que este sea un muón de señal.

Se muestra un ejemplo en la figura 6.7, donde se representan por separado los sucesos que se introducen al modelo como señal y como fondo. Se usará la etiqueta "bad muons" para muones de fondo y "good muon" para muones de señal. Además, se incluyen también los sucesos del entrenamiento. En la figura 6.7, las distribuciones del set entrenamiento y del set de test coinciden a la perfección, lo que representa una situación ideal. En caso del que el set de entrenamiento presente mejores resultados que la del test, es decir, mayor probabilidad de que los muones de señal estén clasificados como señal o los de fondo como fondo, se habría producido sobre entrenamiento del modelo.



Figura 6.7: Ejemplo de distribución de salida de un MVA.

Empleando esta figura como referencia 6.7, ha de determinarse el criterio de selección en el discriminante del MVA. Es decir, a partir de que valor de la probabilidad se tomará el muon como señal. En la decisión de este valor se puede ser más o menos estricto, en función de lo que se pueda permitir en el estudio. En los modelos de este trabajo el criterio que se ha seguido es: el menor valor del discriminante con el que se obtenga una TFP < 0.15, con una precisión en el discriminante de 0.05.

Por último, la figura que se considerará con más peso en la comparación será la representación de la eficiencia del modelo para sucesos en diferentes intervalos de p_T y η . Un ejemplo es la figura 6.8.



Figura 6.8: Ejemplo de presentación de la eficiencia de un MVA de clasificación de muones según en función de p_T y su η .

Es importante estudiar el comportamiento de los MVA en el contexto en el que se van a aplicar. Recordando que los muones de mesones tienen una distribución en p_T muy similar a la del fondo (bajos valores de p_T), mientras que para valores altos de p_T la contribución del fondo es mínima, cabe esperar que la eficiencia del MVA sea menor para valores bajos de p_T y considerablemente alta para valores altos. Tener esto en cuenta ayudará a decidir en qué contexto puede ser útil emplear cada MVA.

Como fuentes de incertidumbre, se tiene en cuenta únicamente la incertidumbre estadística, que se toma como la raíz cuadrada del número de sucesos. Es una incertidumbre poissoniana, ya que el experimento que se está realizando es de conteo.

Capítulo 7

Modelos entrenados

Para el entrenamiento de todos los modelos MVA se ha empleado la biblioteca Scikit-learn de Python, [12], que se ha tomado como lenguaje principal. Como herramienta auxiliar para el análisis se ha empleado ROOT, [4], especialmente para la representación gráfica.

Se han entrenado 5 modelos distintos, que engloban todos los escenarios presentados en el capítulo anterior. El algoritmo de ML del que se ha partido es un Random Forest, con 200 estimadores (árboles de decisión) y 8 niveles de profundidad. La elección de hiperparámetros se realizó tratando de optimizar el entrenamiento para un MVA de RF sin dxy y dz y señal $p_T > 20 \ GeV$, el modelo que se comentará en la sección 7.4. Para permitir la comparación entre modelos, se han mantenido los mismos hiperparámetros en todos los Random Forest.

Al final del estudio se incluye un BDT de 250 estimadores y profundidad 9, que se discutirá en la sección 7.6. En este caso, los hiperparámetros han sido optimizados para el modelo en cuestión.

Según lo descrito en el capítulo 6, cabe suponer que el MVA más inclusivo o general que se podrá construir con la muestra y variables disponibles será un modelo entrenado sobre la muestra con sucesos de $p_T > 10 \text{ GeV}$ sin incluir los parámetros de impacto dxy y dz como variables de entrada. Este es modelo presentado en la sección 7.1. Partiendo de él, se mostrará el resultado de añadir dxy y dz en el entrenamiento (sección 7.2), de desplazar la selección de la muestra a $p_T > 20 \ GeV$ (sección 7.4), de eliminar p_T y η del entrenamiento (sección 7.5) y por último, de emplear un algoritmo de ML diferente (sección 7.6).

7.1. Random Forest sin dxy y dz y señal $p_T > 10 \ GeV$

Como se ha visto en secciones anteriores, se comenzará el estudio con la suposición de que este es el modelo más inclusivo que podemos entrenar con la muestra y variables disponibles. Es decir, el modelo que es más general y se podría usar para más tipos de muones. Se espera que este modelo sea capaz de seleccionar muones de distintas fuentes y en un rango de momentos amplio, desde 10 GeV hasta 100 GeV.

Al tomar todos los sucesos con $p_T > 10 \ GeV$ en la muestra, se tendrán en cuenta tanto los muones de bosones como la contribución de muones de mesones. Cabe recordar que, como mostraba la figura 6.5a, los muones originarios de mesones tienen una distribución en p_T muy similar a la del fondo, por lo que incluirlos en el entrenamiento supondrá una reducción en la eficiencia del modelo. El MVA "confundirá" más sucesos de fondo con muones de este tipo.

Por otro lado, se ha visto que otra diferencia fundamental entre las dos contribuciones de señal se encuentra en los parámetros de impacto. Como los muones de mesones no se forman en el centro del detector, el valor de dxy y dz será mayor. Esta distribución se asemejará más a la del fondo que en el caso de los muones *prompt*. Si se incluyen en el entrenamiento estas variables, el modelo tenderá más a diferenciar entre los dos tipos de muones, y a tomar los de mesones como fondo.

Por lo tanto, este modelo estaría en las mejores condiciones para seleccionar simultáneamente ambas contribuciones de señal, pero veremos que esto empeorará la eficiencia. En la tabla 7.1 se presenta la importancia que el MVA ha dado a cada una de las variables de entrada, su relevancia a la hora de clasificar.

Cabe destacar que la variable de mayor importancia en este caso es la compatibilidad de los segmentos en las cámaras de muones. Además, es interesante que, incluso tras el repesado, la variable p_T tiene cierta importancia en el entrenamiento. Esto querrá indicar que el modelo ha aprendido de las diferencias entre las distribuciones, aunque no la haya empleado directamente para clasificar.

Variables	Importancia
Segment Compatibility	0.27434
$\chi^2 \ local$	0.17213
χ^2 norm	0.15433
Matched stations	0.12104
Valid Fraction	0.07072
Valid hits	0.06607
Tracker Layers	0.04995
Valid Pixel,	0.04749
p_T	0.02902
Global Muon	0.00930
Kink	0.00486
η	0.00076

Tabla 7.1: Importancia de las variables en el MVA: RF sin dxy y dz y señal $p_T > 10 \ GeV$.

En la figura 7.1 se representa la distribución de salida del MVA. Teniendo en cuenta el criterio fijado, se ha determinado la selección en el discriminante en 0.45, para una tasa de falsos positivos de $TFP = 0.1468 \pm 0.0009$. Así, este MVA tendrá una eficiencia total de 0.9594 ± 0.0003 . Se observa que en la figura 7.1 aparece un doble pico de señal. El pico de menor tamaño, alrededor de 0.7, corresponde a la contribución de señal que son esencialmente muones de mesones, mientras que los muones de bosones están en el pico de la derecha del discriminante. Además, en el pico de muones de mesones se presenta también un pico de contaminación de muones de fondo. Tal como se esperaba, el modelo ha errado al diferenciar algunos sucesos de fondo, tomándolos como señal.



Figura 7.1: Distribución de salida del MVA: RF sin dxy y dz y señal $p_T > 10 \ GeV$. Criterio de selección en el discriminante: 0.45.

Otra forma de observar el buen comportamiento del modelo será en las figuras de 7.2. En 7.2a se muestra el conteo de sucesos en cada categoría: señal tomada como seña, señal tomada como fondo, fondo tomado como señal y fondo como fondo. Observamos que, en proporción, una mayor cantidad de sucesos de fondo son mal clasificados. En 7.2b se busca obtener un valor del área bajo la curva cercano a 1, que en este caso puede considerarse bueno. Es de especial interés la comparación del comportamiento del modelo sobre el set de entrenamiento frente al comportamiento sobre el set de test. Como existe poca diferencia, se puede afirmar que el modelo ha aprendido lo suficientemente bien de los datos, sin sobre entrenamiento.



Figura 7.2: Comportamiento del modelo RF sin dxy y dz y señal $p_T > 10 \ GeV$, según:(a) Matriz de confusión, (b) Curva ROC.

Por último, la figura de mayor interés a la hora de comparar los modelos es la siguiente, 7.3. Se observa en ella que a partir de 30 GeV la eficiencia es constante y toma valores por encima del 97 % para todos los rangos de pseudorapidez. A bajo p_T la eficiencia empeora, pero se mantiene alta, por encima del 92 % en todo el rango de p_T . Esta pérdida en eficiencia a bajo p_T es esperable, dado que algunos de los criterios de selección que aplica el MVA serán más difícil de pasar para muones de bajo p_T . Para valores altos de p_T la mayor parte de los sucesos de señal serán de muones prompt, que se diferencia en mayor medida del fondo.

En $|\eta|$, ya se mostró en la figura 6.5b que, como era de esperar, no existe gran diferencia entre las contribuciones de señal. Las diferencias de eficiencia que se observan en función de esta variable pueden deberse a la geometría del detector. Cabe recordar que la zona de barril del detector CMS recorre aproximadamente $0 < |\eta| < 1$, mientras que la de tapones cubre $1.5 < |\eta| <$ 3, por lo que en $1 < |\eta| < 1.5$ se encuentra la zona de transición. En este caso, se observa mayor eficiencia para las zonas de mayor $|\eta|$, siendo más uniforme en la zona de los tapones.



Figura 7.3: Eficiencia en función de p_T y η del modelo RF sin dxy y dz y señal $p_T > 10 \ GeV$.

7.2. Random Forest con todas las variables y señal $p_T > 10 \ GeV$

Se introducen ahora también los parámetros de impacto en el entrenamiento, para comprobar las diferencias que esto genera en la clasificación.

Lo primero que cabe destacar, en la tabla 7.2, es que la variable dz ha pasado a ser la de mayor importancia en el MVA, lo que implica que la diferencia entre sus distribuciones de señal y fondo será de gran relevancia en la clasificación.

Variable	Importancia
dz	0.23747
Segment Compatibility	0.20492
χ^2 norm	0.14770
$\chi^2 \ local$	0.11353
Matched stations	0.09528
Valid hits	0.06004
Valid Fraction	0.04847
Tracker Layers	0.02805
Valid Pixel	0.02558
dxy	0.01366
Global Muon	0.01333
p_T	0.00893
Kink	0.00238
η	0.00066

Tabla 7.2: Importancia de las variables en el MVA: RF con todas las variables y señal $p_T > 10 \ GeV$.

En este caso, se comprobó que la tasa de falsos positivos desciende rápidamente al aumentar el discriminante, por lo que, con el criterio escogido, el primer valor del discriminante que presenta una TFP < 0.15 es 0.2, con un valor de $TFP = 0.0923 \pm 0.0007$. Así, se obtiene una eficiencia total de 0.9866 ± 0.0003 . Cabe resaltar que el criterio de selección en el discriminante ha disminuido bastante en relación al MVA anterior, entrenado sin dxy y dz, resultando además en una mayor eficiencia. Asimismo, en la distribución de salida del MVA, figura 7.4, se observa que ha desaparecido el pico de contaminación de muones de fondo en la señal. Esto puede deberse a que el modelo está usando las variables dz y dxy para clasificar esos muones como fondo mientras que sin dichas variables no era capaz de hacerlo.



Figura 7.4: Distribución de salida del MVA: RF con todas las variables y señal $p_T > 10 \ GeV$. Criterio de selección en el discriminante: 0.2.

Fijándose en la figura 7.5a, se observa de nuevo que el número de sucesos de fondo erróneamente clasificados como señal ha disminuido, que serían los correspondientes al pico de contaminación en la distribución de salida. También se ha reducido el número de muones tomados como fondo. Ambos resultados parecen indicar que incluir dxy y dz ha resultado en una mejora en la clasificación para el total de muones, al contrario de lo esperado, pero sería importante comprobar cómo se están clasificando los muones cada contribución de señal. El modelo ha aprendido de los observables dxy y dz y, a pesar de que para los muones de mesones estas distribuciones se asemejen más a la del fondo que al resto de la señal, no ha supuesto un peor resultado. Como el comportamiento de este modelo se aleja de lo que se había supuesto, será importante estudiar en profundidad como afectan los observables dxy y dz a la clasificación. Se verá en la sección 7.3.



Figura 7.5: Comportamiento del modelo RF con todas las variables y señal $p_T > 10 \ GeV$, según:(a) Matriz de confusión, (b) Curva ROC.

La figura 7.5b muestra que de nuevo el modelo se comporta de forma muy similar para el set de entrenamiento y para el set de test, lo que es un buen indicador para su posible aplicación en distintos sets de datos.



Figura 7.6: Eficiencia en función de p_T y η del modelo RF con todas las variables y señal $p_T>10~{\rm GeV}$

Por último, la figura 7.6 muestra la eficiencia del MVA en función de p_T y η . En este caso se observa que la eficiencia en valores bajos de p_T es considerablemente mejor que en el modelo anterior y mucho más uniforme para todos los valores de η , lo que indica es que las variables dxy y dz ayudan sobre todo a quitar sucesos de fondo de p_T bajo.

Puesto que los resultados parecen contradecir la suposición inicial, se le dedicará una sección a estudiar en mayor profundidad la distribución los parámetros de impacto en los dos modelos anteriores.
7.3. Estudio de la distribución de los parámetros de impacto $dxy \ge dz$

Se van a comparar las distribuciones de dxy y dz para cuatro categorías: los muones de señal clasificados como señal, de señal como fondo, de fondo como señal y de fondo como fondo, para los dos modelos anteriores.



Figura 7.7: Parámetro de impacto dxy en las distribuciones de salida del modelo: (a) RF sin dxy y dz y señal $p_T > 10 \ GeV$, (b) RF con todas las variables y señal $p_T > 10 \ GeV$.

Las diferencias que se observan en la figura 7.7 entre 7.7a y 7.7b en las distribuciones de dxy no son demasiado grandes. En 7.7b no se ve un criterio de selección claro en dxy, lo que indica que el MVA no parece estar aprendiendo nada de esta variable. Las sutiles diferencias que se observan entre 7.7a y 7.7b pueden ser consecuencia de la selección que sí se verá para dz (figura 7.8) ya que ambas variables están bastante correlacionadas.

Ya en la figura 6.6a, se observaban menores diferencias entre las dos contribuciones de señal para la variable dxy. El hecho de que esta variable no cobrase demasiada importancia en el entrenamiento, como se mostraba en 7.2, ya podría indicar que no se produciría un cambio muy considerable en la distribución de cada categoría.

No obstante, para dz se presenta una situación completamente opuesta.

Ya que la variable pasó a tomar el valor más alto de importancia, se espera que se haya modificado considerablemente la distribución de cada categoría al tenerla en cuenta.



Figura 7.8: Parámetro de impacto dz en las distribuciones de salida del modelo: (a) RF sin dxy y dz y señal $p_T > 10 \ GeV$, (b) RF con todas las variables y señal $p_T > 10 \ GeV$.

En la figura 7.8 se ve que la forma de las distribuciones de los muones bien clasificados, señal como señal y fondo como fondo, no han variado demasiado. Esto es de esperar ya que la mayor parte de elementos en estas categorías se mantiene constante de un modelo a otro.

En cambio, han cambiado por completo las distribuciones de los muones mal clasificados. Es importante notar que en 7.8a, correspondiente al modelo que no incluye la variable en el entrenamiento, no se observa un criterio de selección en la distribución en dz donde el MVA clasifique entre señal y fondo. Al no incluirse en el MVA, el modelo no ha aprendido a relacionar los sucesos según esta variable. En la figura 7.8b, por el contrario, sí que se observa claramente que los muones con dz a partir de 0.35 cm son considerados como fondo, mientras que apenas habrá sucesos clasificados como señal. El modelo ha utilizado esta variable en gran medida para la clasificación, por lo que el criterio de selección es drástico.

Aparentemente, el MVA incluyendo todas las variables sí genera un sesgo en dz, como se esperaba. Sin embargo, esta selección estricta sobre la distribución de dz no empeora la eficiencia del modelo, si no que es considerablemente mejor a la del modelo anterior.

Una posible justificación es que el nuevo MVA, que incluye dxy y dz en el entrenamiento, esté mejorando en gran medida la clasificación para los muones de bosones. De esta forma, la tasa de verdaderos positivos aumentaría más, con sucesos de esta contribución de señal, de lo que disminuye al no clasificar correctamente los muones de mesones.

7.4. Random Forest sin $dxy \ge dz \ge ent{and} p_T > 20 \ GeV$

Se continúa el estudio partiendo del modelo original, 7.1, y desplazando el criterio de selección en la muestra a sucesos con $p_T > 20 \text{ GeV}$. Para este modelo sí se espera una mayor eficiencia, especialmente para valores más bajos de p_T , ya que se está eliminando gran parte de los muones de mesones. La señal estará en mayor proporción compuesta de muones *prompt*, lo que facilitará la selección.

Variables	Importancia	
Segment Compatibility	0.35865	
χ^2 norm	0.18347	
Matched stations	0.16640	
$\chi^2 \ local$	0.12142	
Valid hits	0.06745	
Valid Fraction	0.03510	
Tracker Layers	0.03110	
Valid Pixel	0.02050	
p_T	0.00704	
Global Muon	0.00374	
Kink	0.00208	
η	0.00095	

Tabla 7.3: Importancia de las variables en el MVA: RF sin dxyydzy señal $p_T>20\;GeV$

La importancia de las variables, en la tabla 7.3, no ha variado significativamente de la del modelo original. Ha cobrado importancia la variable *Matched stations*, que hace referencia al número de estaciones en las cámaras de muones con señales. Generalmente, tan sólo los muones llegan hasta las cámaras de muones y, en particular, los de alto p_T dejarán un mayor número de trazas. Así, es lógico que al eliminar los muones de bajo p_T se acentúe la relevancia de esta variable en el entrenamiento.

Es interesante observar también que, a pesar de eliminar los sucesos de bajo p_T , dejando una distribución en p_T mucho más diferenciada para señal y fondo, esta variable no cobra mayor importancia en el entrenamiento. Este es el resultado esperado del repesado de la variable.



Figura 7.9: Distribución de salida del MVA: RF sin dxy y dz y señal $p_T > 20 \ GeV$. Criterio de selección en el discriminante: 0.05.

La distribución de salida del MVA, que se puede ver en la figura 7.9, muestra en este caso muestra una distinción muy clara entre los sucesos de fondo y señal. De hecho, manteniendo el criterio anterior (TFP < 0.15), el criterio de selección en el discriminante de este modelo se fija en 0.05, con $TFP = 0.1362 \pm 0.0016$ y con una eficiencia total de 0.9975 ± 0.0004 .



Figura 7.10: Comportamiento del modelo RF sin dxy y dz y señal $p_T > 20 \ GeV$, según:(a) Matriz de confusión, (b) Curva ROC.

En la figura 7.10 se ve de nuevo que, para la muestra tomada, el modelo funciona muy correctamente. En 7.10a se muestra que la tasa de falsos positivos es extremadamente baja, y la de falsos positivos es también la menor hasta el momento. En 7.10b se ve también que el área bajo la curva ROC muy próxima a uno, e igual para el test y train.

Para terminar, se observa en la figura 7.11 que este modelo tiene una eficiencia muy alta para cualquier valor de p_T y en todo η . Es importante recordar que este modelo no es directamente comparable con el modelo inicial, si no que este está mejor optimizado para la selección de muones *prompt*, que son la gran mayoría de los que se incluyen en el entrenamiento. Sin embargo, este modelo sería muy propenso a no seleccionar los muones de bajo p_T , incluso los *prompt*. Por lo tanto, aunque su respuesta parece muy precisa, habría que examinar cuidadosamente en que escenarios utilizarlo. Esto es, como está entrenado con sucesos de $p_T > 20 \ GeV$ no está garantizado que funcione con la misma eficiencia a p_T más bajo, por tanto no sería un modelo ideal si se quisiera usarlo para seleccionar muones de $p_T < 20 \ GeV$.



Figura 7.11: Eficiencia en función de p_T y η del modelo RF sin dxy y dz y señal $p_T > 20 \ GeV$.

7.5. Random Forest sin dxy y dz, p_T y η , y señal $p_T > 10 \ GeV$

El objetivo de este modelo es principalmente comprobar si el repesado de las variables p_T y η está influyendo en el modelo como se pretendía. Se espera un modelo cuyo comportamiento no difiera en gran medida el modelo inicial, 7.1, aunque sí con una eficiencia algo peor.

Comprobamos en la tabla 7.4 que la importancia de las variables no se ha modificado significativamente de 7.1. Aunque no siguen exactamente el mismo orden, la relación entre las variables más relevantes para la clasificación y las de menor importancia es muy similar, excluyendo en este caso las variables eliminadas, p_T y η .

Variables	Importancia	
Segment Compatibility	0.28453	
χ^2 norm	0.20886	
Matched stations	0.14445	
$\chi^2 \ local$	0.13216	
Valid hits	0.06351	
Valid Fraction	0.06122	
Valid Pixel	0.05365	
Tracker Layers	0.03746	
Global Muon	0.01073	
Kink	0.00343	

Tabla 7.4: Importancia de las variables en el MVA: RF sin dxy, dz, $p_T \neq \eta$, y señal $p_T > 10 \ GeV$.

En la distribución de salida del MVA, figura 7.12, se mantiene el mismo criterio de selección el discriminante: 0.45, con una eficiencia total de 0.9503 ± 0.0003 , similar a la del modelo inicial, en este caso para una $TFP = 0.1478 \pm 0.0010$ En la figura se puede notar que, en los modelos de señal, se ha reducido la probabilidad de ser clasificados como buenos, pero con la elección manual del discriminante esto no afecta a la eficiencia del modelo. Al contrario que en el modelo inicial, 7.7a, aquí no se observa un doble pico en la distribución de señal. Al no introducir la variable p_T no se genera la misma distinción entre los muones de mesones y de bosones. No obstante, sí se encuentra un pico de contaminación de sucesos de fondo sobre la distribución de señal.

En las figuras 7.13 sí que se puede observar que el comportamiento del modelo es ligeramente peor que en el inicial. En 7.13a, el número de sucesos mal clasificados es ligeramente mayor y en 7.13a, el área bajo la curva ROC ha disminuido también.

En la representación de la eficiencia en función de p_T y η aparece la mayor prueba de diferencia hasta el momento. Al contrario que en todos los modelos anteriores, la eficiencia parece decaer para valores altos de p_T . Esto se debe a que, aunque para valores bajos de p_T este no era necesariamente el mejor discriminante, para valores altos existía una diferencia muy considerable ente su distribución de señal y fondo. Al eliminar la variable del modelo, se pierde la información obtenida de la correlación que hay entre p_T y variables



Figura 7.12: Distribución de salida del MVA: RF sin dxy, dz, $p_T \ge \eta$, $\ge n_T \ge 10$ GeV. Criterio de selección en el discriminante: 0.45.



Figura 7.13: Comportamiento del modelo RF sin dxy, dz, p_T y η , y señal $p_T > 10 \ GeV$, según: (a) Matriz de confusión, (b) Curva ROC.

como, por ejemplo, número de estaciones con señal, que es relevante para la clasificación.



Figura 7.14: Eficiencia en función de p_T y η del modelo RF sin dxy, dz, p_T y η y señal $p_T > 10 \ GeV$.

En definitiva, se puede afirmar que el repesado de p_T y η cumple con su objetivo: las variables no se emplean como discriminante principal, 7.1, pero el modelo aprende lo suficiente de la forma de sus distribuciones y la correlación con el resto de las variables.

7.6. Boosted Decision Trees sin dxy y dz y señal $p_T > 10 \ GeV$

El último modelo que se presentará en este trabajo es un modelo equivalente al inicial, en cuanto a parámetros físicos, pero empleando un algoritmo de ML diferente. En todos los MVAs anteriores se empleó un algoritmo de Random Forest que, como se explicó en el capítulo 5, se basa en un método de *bagging*. En este caso, para explorar sus diferencias, se empleará un BDT, que emplea el método *boosting*.

Variables	Importancia	
η	0.20235	
p_T	0.14567	
$\chi^2 \ local$	0.13381	
Segment Compatibility	0.11922	
Valid hits	0.09089	
χ^2 norm	0.08909	
Kink	0.06333	
Valid Pixel	0.04268	
Tracker Layers	0.04124	
Valid Fraction	0.03971	
Matched stations	0.02917	
Global Muon	0.00283	

Tabla 7.5: Importancia de las variables en el MVA: BDT sin dxy y dz y señal $p_T > 10 \ GeV$.

Ambos algoritmos funcionan de manera muy diferente, y la primera evidencia clara se muestra en la tabla 7.5. Mientras que en todos los casos anteriores se encontraban hacia al final de la tabla, η y p_T se toman como las variables de entrada de mayor importancia, incluso aunque las estemos introduciendo como repesadas en el entrenamiento. En este caso, como ya se ha dicho, la diferencia no corresponde al contexto físico para el que se entrena el MVA, si no únicamente al funcionamiento interno del algoritmo.

En la figura 7.15 se observa una distribución de salida mucho más diferenciada entre señal y fondo, sin el pico de contaminación que se encontraba en el modelo inicial. Aun así, el criterio de selección en el discriminante no se ha desplazado demasiado: se fija en 0.4 con una eficiencia total de 0.9761 ± 0.0003 , para una $TFP = 0.1432 \pm 0.0009$.



Figura 7.15: Distribución de salida del MVA: BDT sin dxy y dz y señal $p_T > 10 \ GeV$. Criterio de selección en el discriminante: 0.40.

En la figura 7.16a encontramos de nuevo una tasa más baja de falsos positivos y de falsos negativos, mientras que en 7.16b se muestra un área bajo la curva más cercano a 1 al que se encontraba en 7.2b.



Figura 7.16: Comportamiento del modelo BDT sin dxy y dz y señal $p_T > 10 \ GeV$, según: (a) Matriz de confusión, (b) Curva ROC.

Por último, examinando la representación de la eficiencia en función de p_T y η , se observa que sigue una distribución muy similar a la del modelo inicial, pero más alta en todo caso.

En general, el MVA entrenado con este algoritmo parece tener un mejor resultado que el entrenado con el algoritmo Random Forest. Aun así, al obtenerse resultados similares parece que el Random Forest y el y el BDT son compatibles, por lo que se puede decir que son modelos robustos.



Figura 7.17: Eficiencia en función de p_T y η del modelo XGB sin dxy y dz y señal $p_T > 10 \ GeV$.

7.7. Comparación de los modelos

Para finalizar el capítulo se recogerá en esta sección un resumen de todos los modelos presentados, con sus características fundamentales y las diferencias que presentan con el modelo que se ha tomado como referencia. En la tabla 7.6 se presentan los MVA entrenados, con el criterio de selección que se empleó en su discriminante, la TFP fijada y la eficiencia que se obtiene con ellos.

MVA	Criterio	TFP	Eficiencia
RF sin dxy y dz ($p_T > 10 \ GeV$)	0.45	0.1468 ± 0.0009	0.9594 ± 0.0003
RF todas las variables $(p_T > 10 \ GeV)$	0.20	0.0923 ± 0.0007	0.9866 ± 0.0003
RF sin dxy y dz ($p_T > 20 \ GeV$)	0.05	0.1362 ± 0.0016	0.9975 ± 0.0004
RF sin dxy , dz , $\eta \neq p_T$ ($p_T > 10 \ GeV$)	0.45	0.1478 ± 0.0010	0.9503 ± 0.0003
BDT sin dxy y dz ($p_T > 10 \ GeV$)	0.40	0.1432 ± 0.0009	0.9761 ± 0.0003

Tabla 7.6: Resumen de los modelos entrenados con el criterio de selección en su discriminante, la tasa de falsos positivos y el valor de la eficiencia.

Cabe recordar que el modelo del que se parte es un RF entrenado con sucesos de $p_T > 10 \ GeV$ sin dxy y dz, ya que se supuso inicialmente que estas variables podrían generar un sesgo para los muones de mesones. El resultado de este modelo es bastante satisfactorio, con alta eficiencia, aunque se observa una distinción entre la clasificación de muones *prompt* y muones de mesones. Estos últimos tendrán menor probabilidad de clasificarse correctamente. Tras comparar el modelo inicial con el entrenado en mismas condiciones, pero incluyendo todas las variables de entrada, se observó que, en efecto, se generaba un sesgo en la clasificación de seña según dz. A partir de $dz = 0.35 \ cm$, todos los sucesos se clasificaron como fondo, reduciendo así la identificación de muones de bosones. Aun así, la eficiencia del modelo se vio incrementada, es decir, la información añadida mejoró la clasificación lo suficiente como para compensar el sesgo en dz.

A continuación, se comparó el modelo inicial con uno entrenado tan sólo incluyendo sucesos de $p_T > 20 \ GeV$ en la muestra. Esta modificación pretende excluir del entrenamiento la mayor parte de los muones de mesones, homogeneizando las características físicas de los sucesos de señal. Tal como se esperaba, este modelo resulta en una mayor eficiencia. Sin embargo, al haber aprendido tan sólo a clasificar muones *prompt*, no será tan eficaz si se emplea sobre una muestra más variada.

El siguiente caso a comparar es el modelo inicial, RF con señal de $p_T > 10 \ GeV$ sin dxy y dz, frente al modelo entrenado eliminando también del entrenamiento p_T y η . El objetivo en este caso era comprobar cómo afecta al MVA el repesado de p_T y η . Se comprueba que el resultado es el esperado, el modelo no ha cambiado demasiado. Se mantiene la relación entre las variables y una eficiencia alta. Sin embargo, ha desaparecido la distinción que se

encontraba en el modelo inicial entre el discriminante de los muones *prompt* y los muones de mesones, indicando que esta se aprendió de la variable p_T en el primer modelo.

Por último, se ha entrenado un modelo en las condiciones del modelo inicial, señal de $p_T > 10 \ GeV \sin dxy$ y dz, pero utilizando un algoritmo de boosting (BDT), en lugar del RF. El resultado en este caso es incluso mejor que para el modelo inicial, con una eficiencia más alta para todo valor de p_T . Obtener resultados similares usando dos modelos diferentes como son un Random Forest y un BDT muestra que los modelos entrenados son robustos.

Conclusiones

En este trabajo se han presentado las razones por las que es imprescindible una buena identificación de leptones en los sucesos registrados por el experimento CMS del LHC. Cualquier análisis de datos tomados por CMS necesita de una reconstrucción e identificación precisa de las partículas del estado final, lo que permitirá el estudio de distintos procesos físicos, entre otros, medidas del modelo estándar, búsquedas de partículas exóticas o de materia oscura. Para el estudio de muchos de estos procesos, la identificación de leptones y, en particular, de muones, es de vital importancia. Además, debe tenerse en cuenta que existen diversas fuentes de muones (en general, provenientes de bosones o provenientes de mesones ligeros o B), que presentas distintas características físicas. Por ello, el proceso de identificación de cada tipo de muones puede ser diferente.

Se han introducido diferentes técnicas de aprendizaje automático que podrían resultar en una mejora importante en la tarea de selección de muones. Este trabajo se realiza justo antes del comienzo del Run 3, por tanto, tiene sentido mejorar la identificación en preparación para la llegada de nuevos datos que analizar.

En concreto, se han entrenado cinco modelos con diversas variables de entrada y contribución de señal en el entrenamiento, con intención de comparar cómo afectan estas a la selección de muones *prompt*, que provienen de bosones W o Z, y de muones de mesones. Se incluyen dos algoritmos de Machine Learning: Random Forest y Boosted Decision Trees. En definitiva, se han comparado cuatro características del diseño del MVA.

Primero, se ha considerado si es conveniente incluir los parámetros de

impacto en el entrenamiento, a causa de la diferencia que presentan estos observables según las dos contribuciones de señal. Tras la comparación, se ha observado que el modelo que incluye todas las variables obtiene mejores resultados en la clasificación que el que no incluye dxy y dz. Se ha visto que el MVA da bastante importancia a estas distribuciones y que sí aparece un sesgo que resulte en la exclusión de los muones de mesones en la clasificación. No obstante, el modelo parece mejorar lo suficiente la selección de muones de bosones como para subsanar la pérdida en eficiencia que podría generar el sesgo. Es decir, aunque mejora la selección para el total de muones, el modelo está optimizado para la selección de muones de bosones y no es recomendable para la selección de muones.

A continuación, se ha comparado el resultado de restringir la muestra a sucesos de $p_T > 20 \ GeV$, excluyendo gran parte de los muones de mesones, frente a incluir todos los sucesos de $p_T > 10 \ GeV$. El MVA con la muestra de p_T más altos presenta un mejor resultado, ya que los sucesos de señal tienen todos características físicas muy similares, por lo que es más fácil distinguirlas del fondo. Este modelo, no obstante, está optimizado para la selección de muones *prompt* y no sería ideal para muestras más variadas.

Seguidamente, se comprueba la eficacia del repesado de los observable p_T y η como variables de entrada en el MVA. El objetivo del repesado era poder introducirlas en el entrenamiento sin que supusieran un sesgo entre las contribuciones de señal. La conclusión en este caso es que, en efecto, el repesado de las variables consigue que el modelo aprenda suficientemente de sus distribuciones, sin modificar sustancialmente los criterios de selección.

Por último, se compara el resultado del MVA entrenado mediante un algoritmo de *bagging* frente a un algoritmo de *boosting*. Se ha podido observar, por la importancia que toman las variables, que el funcionamiento interno de los algoritmos es completamente distinto. Aunque ambos modelos devuelven buenos resultados, el MVA de boosting tiene una eficiencia más alta que el de bagging, en todo el rango de p_T y η , por lo que puede considerarse ligeramente superior.

A pesar de las diferencias que presentan, todos los modelos de aprendizaje automático presentados tienen una eficiencia alta, en torno al 90% y manteniendo una tasa de falsos positivos baja, menor que 0.15. Además, cuentan con la ventaja de poder adaptarlos al contexto físico al que se quieren aplicar ajustando correctamente las variables y muestra de entrenamiento.

El desarrollo de técnicas de inteligencia artificial, como las presentadas en este trabajo, será crucial para optimizar la identificación de muones en el experimento CMS. Un mejor tratamiento de datos en el LHC facilitará enormemente el estudio de gran cantidad de procesos físicos de interés y dará pie al posible descubrimiento de nueva física.

Una continuación natural de este estudio sería escoger uno de los modelos entrenados para su implementación en la identificación de muones de CMS y comprobar si funciona bien en datos reales, cuando se tengan los primeros datos del Run 3 que comienza este julio.

Bibiografía

- [1] GL Bayatian y col. CMS Physics: Technical Design Report Volume 1: Detector Performance and Software. Inf. téc. CMS-TDR-008-1, 2006.
- [2] Florian Beaudette. "FAMOS, a FAst MOnte-Carlo Simulation for CMS". En: (2005).
- [3] Javier Brochero. "Measurement of the inclusive ttbar cross section in the LHC". En: arXiv preprint arXiv:1412.7176 (2014).
- [4] Rene Brun y Fons Rademakers. "ROOT An Object Oriented Data Analysis Framework". En: Nucl. Inst. Meth. in Phys. Res. A 389 (1997), págs. 81-86.
- [5] CERN. The Large Hadron Collider. URL: https://home.cern/science/ accelerators/large-hadron-collider.
- [6] CMS collaboration, CMS Collaboration y col. "Particle-flow event reconstruction in CMS and performance for jets, taus and MET". En: (2009).
- [7] Abhigyan Dasgupta. Study of neutron-induced background hits in the CMS endcap muon system. Inf. téc. 2017.
- [8] Siona Ruth Davis. "Interactive Slice of the CMS detector". En: (ago. de 2016). URL: https://cds.cern.ch/record/2205172.
- [9] Aurélien Géron. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. "O'Reilly Media, Inc.", 2019.
- [10] Sam Kash Kachigan. Multivariate statistical analysis: A conceptual introduction. Radius Press, 1991.
- [11] Martijn Mulders. "Muon Reconstruction and Identification at CMS". En: Nucl. Phys. B, Proc. Suppl. 172 (dic. de 2006), 205-207. 4 p. DOI: 10.1016/j.nuclphysbps.2007.08.049. URL: https://cds.cern. ch/record/1009071.

- [12] F. Pedregosa y col. "Scikit-learn: Machine Learning in Python". En: Journal of Machine Learning Research 12 (2011), págs. 2825-2830.
- [13] Ana Porras. What is the difference between bagging and boosting? Quantdare. Nov. de 2020. URL: https://quantdare.com/what-is-thedifference-between-bagging-and-boosting/.
- [14] Bogdan Povh y col. *Particles and nuclei*. Fifth. Springer, 2006.
- [15] Albert M Sirunyan, CMS Collaboration y col. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at s = 13 TeV". En: (2018).
- [16] Albert M Sirunyan y col. "Particle-flow reconstruction and global event description with the CMS detector". En: Journal on Instrumentation (2017).
- [17] The standard model. URL: https://home.cern/science/physics/ standard-model.
- [18] César Tomé. *Del Modelo Estándar*. Jun. de 2014. URL: https:// culturacientifica.com/2014/06/17/del-modelo-estandar/.
- [19] P.A. Zyla y col. "Review of Particle Physics". En: *PTEP* 2020.8 (2020).
 and 2021 update, pág. 083C01. DOI: 10.1093/ptep/ptaa104.