

UNIVERSIDAD NACIONAL AUTÓNOMA DE
HONDURAS



APLICACIÓN DE ALGORITMOS DE
MACHINE LEARNING PARA LA
IDENTIFICACIÓN DE PARTÍCULAS EN EL
EXPERIMENTO LHCb



TESIS

PARA OBTENER EL TÍTULO DE

MÁSTER EN MATEMÁTICA CON ORIENTACIÓN EN
ESTADÍSTICA MATEMÁTICA

PRESENTA

EYVILIN WALDINA VELÁSQUEZ SÁNCHEZ

ASESORA: PHD. MELISSA MARÍA CRUZ TORRES

TEGUCIGALPA, M.D.C

2022

DEDICATORIA

En memoria de mi padre, aunque no este presente físicamente, su amor
seguirá conmigo todos los días de mi vida.

Agradecimientos

En primer lugar quiero agradecer a Dios por brindarme salud, fortaleza, por ser mi guía y acompañarme en el transcurso de mi vida. A mi madre Mirian Sánchez, que es el motor que me impulsa a seguir adelante, le doy gracias por su esfuerzo, dedicación y apoyo incondicional, que me han llevado donde ahora estoy. A mis hermanos Walter Velásquez y Keren Velásquez, por su cariño y estar conmigo en todo momento. Con mucho amor quiero agradecer a mi novio, por darme ánimos en los momentos más difíciles de esta tesis, por acompañarme en las noches de desvelos y por siempre creer en mí. A toda mi familia, por sus oraciones y palabras de aliento, se que siempre estarán ahí para mí.

Quiero darle las gracias infinitamente a mi asesora PhD. Melissa Cruz, ninguna palabra puede expresar lo agradecida que estoy con usted. Gracias por tener la paciencia y buena voluntad de explicarme todo desde lo más simple hasta lo más abstracto. Gracias por haberme guiado en todo este proceso respondiendo cada una de mis dudas. Gracias por ser una parte fundamental en este logro de mi vida.

Quiero agradecer a mis amigas (Sonia, Delmis y Bessy), por darme siempre palabras de aliento y estar presentes en cada etapa de mi vida.

Resumen

La presente disertación tiene como objetivo la aplicación de algoritmos de machine learning basados en técnicas de análisis multivariante para la reducción del ruido en los canales $B^+ \rightarrow h^+ h'^+ h''^-$ particularmente los canales $B^+ \rightarrow \pi^+ \pi^+ \pi^-$ y $B^+ \rightarrow K^+ K^- \pi^+$ que son los estudiados en esta disertación. Se utilizó parte de la muestra de datos recolectadas por el experimento LHCb en el Run 2, los cuales comprenden los años 2015 y 2016, con una luminosidad integrada de 1.9 fb^{-1} . La determinación de un adecuado criterio de selección es fundamental para todo análisis en donde se extraen medidas Físicas a partir de grandes cantidades de datos. La correcta selección de eventos representa un desafío, pues, usualmente no únicamente están presentes eventos de señal sino que también eventos de ruido de distintas fuentes.

En los canales $B^+ \rightarrow \pi^+ \pi^+ \pi^-$ y $B^+ \rightarrow K^+ K^- \pi^+$ el ruido denominado como combinatorial es reducido considerablemente utilizando Boosted Decision Trees. En una segunda etapa el ruido de mis-identificación es reducido. Esta disertación presenta en detalle los estudios realizados para la determinación de este criterio de selección a través de herramientas estadísticas y algoritmos de machine learning. La eficiencia de la señal es de 88.09% y 82.85% para $B^+ \rightarrow \pi^+ \pi^+ \pi^-$ y $B^+ \rightarrow K^+ K^- \pi^+$, respectivamente. Cabe resaltar que para ambos canales en el corte de Boosted decision trees se alcanza la máxima significancia estadística y la máxima pureza.

Palabras Clave: Análisis Multivariante (MVA), Boosted Decision Trees (BDT), Machine Learning (ML), Decaimientos del mesón B

Abstract

The objective of this dissertation is the application of machine learning algorithms based on multivariate analysis techniques for the reduction of background in the channels $B^+ \rightarrow h^+ h'^+ h''^-$ particularly the channels $B^+ \rightarrow \pi^+ \pi^+ \pi^-$ and $B^+ \rightarrow K^+ K^- \pi^+$ which are the ones studied in this dissertation. Part of the large datasets sample collected by the LHCb experiment in the Run 2 was used, which includes the years 2015 and 2016, with an integrated luminosity of 1.9 fb^{-1} . The determination of an adequate selection criterion is essential for any analysis where Physical measures are extracted from large amounts of data. The correct selection of events represents a challenge, since not only signal events are usually present, but also noise events from different sources.

In the channels $B^+ \rightarrow \pi^+ \pi^+ \pi^-$ and $B^+ \rightarrow K^+ K^- \pi^+$ the background denominated as combinatorial is considerably reduced using Boosted Decision Trees. In a second stage the noise of mis-identification is reduced. This dissertation presents in detail the studies carried out to determine this selection criterion through statistical tools and machine learning algorithms. The signal efficiency is 88.09% and 82.85% for $B^+ \rightarrow \pi^+ \pi^+ \pi^-$ and $B^+ \rightarrow K^+ K^- \pi^+$, respectively. It should be noted that for both channels, the cut of Boosted decision trees the maximum statistical significance and the maximum purity are reached.

Keywords: Multivariate Analysis (MVA), Boosted Decision Trees (BDT), Machine Learning (ML), B -meson decays

Índice general

Índice de Figuras	VIII
Índice de Tablas	XIII
1. Introducción	1
2. Fundamentos Teóricos	4
2.1. Motivación	4
2.2. Modelo Estándar	6
2.3. Conceptos Estadísticos	9
2.3.1. Distribuciones de probabilidad	9
2.3.2. Pruebas de hipótesis	16
2.3.3. El Lema de Neyman-Pearson	19
2.3.4. Razón de Verosimilitud Proyectiva Discriminante . .	20
3. Experimento LHCb	22
3.1. El Gran Colisionador de Hadrones	22
3.2. Datos colectados en el LHC	26
3.3. El Detector LHCb	27
3.4. Subdetectores del LHCb	29
3.4.1. Vertex Locator (VELO)	30
3.4.2. El Imán	31
3.4.3. Sistema de rastreo	32
3.4.4. Detectores Cherenkov (RICH) de imágenes anulares	36
3.4.5. Calorímetros	38
3.4.6. Sistema de muones	39

3.5.	Sistema de Trigger	40
3.5.1.	Nivel-0 (L0)	42
3.5.2.	Trigger de alto nivel (HLT)	43
3.6.	Sistemas de software	44
4.	Análisis de los decaimientos $B^+ \rightarrow \pi^+\pi^+\pi^-$ y $B^+ \rightarrow K^+K^-\pi^+$	46
4.1.	Conjunto de Datos y Simulación	46
4.2.	Variables de interés	47
4.3.	Selección	49
4.3.1.	Proceso de Selección	50
4.4.	Distribución del Mesón B	54
4.4.1.	Tipos de ruido	55
4.5.	Selección Offline	58
4.5.1.	Selección Análisis Multivariante	58
4.5.2.	Selección de Identificación de Partículas (PID)	58
4.6.	Optimización	59
5.	Machine Learning	61
5.1.	Machine Learning en la Ciencia de Datos	61
5.2.	Clasificación en Machine Learning	63
5.3.	Motivación para utilizar métodos multivariantes (MVA)	67
5.4.	Algoritmos de Machine Learning	69
5.4.1.	Boosted Decision Trees (BDT)	69
5.4.2.	Análisis Discriminante lineal (LD)	75
5.4.3.	Estimación de Verosimilitud Proyectiva	76
5.5.	TMVA: Evaluación del desempeño de la clasificación	77
5.5.1.	Curvas ROC	78
6.	Análisis y Resultados	80
6.1.	Estrategia de análisis	80
6.2.	Selección Análisis Multivariante	81
6.2.1.	Entrenamiento de BDT	81
6.2.2.	Optimización de BDT	88
6.2.3.	Aplicación del corte óptimo de BDT	96
6.3.	Selección de Identificación de Partículas (PID)	96
6.3.1.	Optimización (PID)	96

6.3.2. Aplicación del criterio de selección PID	102
7. Conclusiones y Trabajos Futuros	105
7.1. Conclusiones	105
7.2. Trabajos Futuros	106
A. Análisis Multivariante	108
A.1. Separación de las variables	108
A.2. Overtraining	110
B. Curvas ROC - Identificación de Partículas	112
Bibliografía	115

Índice de Figuras

2.1. Grandes asimetrías en regiones del espacio de fase	5
2.2. Partículas elementales e interacciones fundamentales en la naturaleza	8
2.3. Región crítica de una prueba estadística (Tomado de [18]) .	17
2.4. Distribuciones de probabilidad (pdf) del estadístico de prueba escalar $y(x_1, \dots, x_n)$ bajo las hipótesis de señal y ruido. .	18
2.5. Curva ROC: Eficiencia de la señal frente a la probabilidad de mis-identificación de ruido.	19
3.1. CERN	23
3.2. Vista esquemática del complejo LHC, que muestra los diferentes puntos de interacción de sus cuatro experimentos principales	24
3.3. Principales experimentos del LHC	25
3.4. La luminosidad nos da el número de colisiones posibles por unidad de área y tiempo.	26
3.5. Luminosidad integrada producida por el LHC 2010 - 2018 .	27
3.6. Vista lateral del detector LHCb y sus subdetectores	28
3.7. En colisiones de alta energía, simulación de Monte Carlo que muestra la producción de pares $b\bar{b}$ en ángulos pequeños. . .	29
3.8. Imágenes que muestran una vista exterior del módulo VELO en el LHCb	30
3.9. Representación del sensor $\mathcal{R}-$ y del sensor $\varphi-$	31
3.10. Imán dipolo del experimento LHCb	32

3.11. Representación de las estaciones de rastreo del detector LHCb. En la figura están presentes las estaciones de rastreo (púrpura) y el rastreador externo (verde).	33
3.12. La disposición “ $x - u - v - x$ ” del TT, y las cuatro capas de las estaciones.	34
3.13. Vista frontal de los módulos detectores, que rodea el haz de luz central del LHC.	35
3.14. Visualización del paso de una partícula cargada a través de un módulo OT	36
3.15. Representación esquemática de los subdetectores (izquierda) RICH-1 y (derecha) RICH-2	37
3.16. Representación esquemática	39
3.17. Vista lateral del sistema de muones de M1-M5. También se muestran las cuatro regiones R1-R4	41
3.18. Representación de los niveles del Sistema Trigger (Imagen tomada de [17])	42
3.19. Representación de los sistemas de software de procesamiento de datos del LHCb, [24]	44
4.1. Topología de las variables	48
4.2. Proceso de Selección	50
4.3. Distribución de probabilidad de, (a): d1_ProbNNpi, (b): d2_ProbNNpi y (c): d3_ProbNNpi de $B^+ \rightarrow \pi^+ \pi^+ \pi^-$	52
4.4. Distribución de probabilidad de, (a): d1_ProbNNK, (b): d2_ProbNNpi y (c): d3_ProbNNK de $B^+ \rightarrow K^+ K^- \pi^+$	53
4.5. Distribución del Mesón B, canal $B^+ \rightarrow K^+ K^- \pi^+$	55
4.6. Distribución del Mesón B: área roja ($B_m > 5400$) ruido combinatorial, área verde: señal mas ruido de distintos tipos, área azul: predominante los decaimientos parcialmente reconstruidos de cuatro cuerpos.	57
5.1. Algoritmo de machine learning	63
5.2. Estructura de un árbol de decisión, tomado de [3]	70
5.3. Ejemplo del rechazo de ruido vs eficiencia de la señal (“curva ROC”) obtenida cortando las salidas del clasificador para los eventos de la muestra de prueba. Imagen tomada de [3]	79

5.4.	Determinación del mejor clasificador mediante las áreas bajo la curva ROC	79
6.1.	Estrategia de análisis	81
6.2.	Distribuciones de probabilidad (PDFs) de las variables de entrada para el entrenamiento del análisis multivariante, para el canal $B^+ \rightarrow \pi^+\pi^+\pi^-$. La señal se representa en azul y el ruido en rojo.	83
6.3.	Distribuciones de probabilidad (PDFs) de las variables de entrada para el entrenamiento del análisis multivariante, para el canal $B^+ \rightarrow K^+K^-\pi^+$. La señal se representa en azul y el ruido en rojo.	84
6.4.	Matriz de correlación de ruido para el canal $B^+ \rightarrow \pi^+\pi^+\pi^-$ de las 10 variables seleccionadas para el entrenamiento. . .	85
6.5.	Matriz de correlación de señal para el canal $B^+ \rightarrow \pi^+\pi^+\pi^-$ de las 10 variables seleccionadas para el entrenamiento. . .	85
6.6.	Matriz de correlación de ruido para el canal $B^+ \rightarrow K^+K^-\pi^+$ de las 10 variables seleccionadas para el entrenamiento. . .	86
6.7.	Matriz de correlación de señal para el canal $B^+ \rightarrow K^+K^-\pi^+$ de las 10 variables seleccionadas para el entrenamiento. . .	86
6.8.	Curvas ROC-canal $B^+ \rightarrow \pi^+\pi^+\pi^-$: Eficiencia de la Señal frente al Rechazo de Ruido.	87
6.9.	Curvas ROC- $B^+ \rightarrow K^+K^-\pi^+$: Eficiencia de la Señal frente al Rechazo de Ruido.	88
6.10.	Distribución de probabilidad de BDT - $B^+ \rightarrow \pi^+\pi^+\pi^-$ año 2015 y 2016.	89
6.11.	Distribución de probabilidad de BDT - $B^+ \rightarrow K^+K^-\pi^+$ año 2015 y 2016.	89
6.12.	Eficiencias en Función del Corte aplicado en la Salida de BDT, canal $B^+ \rightarrow \pi^+\pi^+\pi^-$	90
6.13.	Eficiencias en Función del Corte aplicado en la Salida de BDT, $B^+ \rightarrow K^+K^-\pi^+$	91
6.14.	Figura de Merito (FoM) de la significancia estadística canal $B^+ \rightarrow \pi^+\pi^+\pi^-$	92
6.15.	Eficiencia de la señal frente al rechazo de background canal $B^+ \rightarrow \pi^+\pi^+\pi^-$	93

6.16. En la figura (a): FoM de la eficiencia. (b): FoM pureza . . .	93
6.17. FoM de la significancia estadística canal $B^+ \rightarrow K^+ K^- \pi^+$.	94
6.18. Eficiencia de la señal frente al rechazo de background canal $B^+ \rightarrow K^+ K^- \pi^+$	95
6.19. En la figura (a): FoM de la eficiencia. (b): FoM pureza . . .	95
6.20. Distribución de Probabilidad de B_m (arriba) y Distribución de Probabilidad de B_m con el corte de BDT (abajo), canal $B^+ \rightarrow \pi^+ \pi^+ \pi^-$	97
6.21. Distribución de Probabilidad de B_m (arriba) y Distribución de Probabilidad de B_m con el corte de BDT (abajo), canal $B^+ \rightarrow K^+ K^- \pi^+$	98
6.22. Izquierda: FoM significancia estadística, Derecha: FoM Efi- ciencia, indicando el corte PID para d1_ProbNNpi de $B^+ \rightarrow$ $\pi^+ \pi^+ \pi^-$	99
6.23. Izquierda: FoM significancia estadística, Derecha: FoM Efi- ciencia, indicando el corte PID para d2_ProbNNpi de $B^+ \rightarrow$ $\pi^+ \pi^+ \pi^-$	99
6.24. Izquierda: FoM significancia estadística, Derecha: FoM Efi- ciencia, indicando el corte PID para d3_ProbNNpi de $B^+ \rightarrow$ $\pi^+ \pi^+ \pi^-$	100
6.25. Izquierda: FoM significancia estadística, Derecha: FoM Efi- ciencia, indicando el corte PID para d1_ProbNNK de $B^+ \rightarrow$ $K^+ K^- \pi^+$	101
6.26. Izquierda: FoM significancia estadística, Derecha: FoM Efi- ciencia, indicando el corte PID para d2_ProbNNpi de $B^+ \rightarrow$ $K^+ K^- \pi^+$	101
6.27. Izquierda: FoM significancia estadística, Derecha: FoM Efi- ciencia, indicando el corte PID para d3_ProbNNK de $B^+ \rightarrow$ $K^+ K^- \pi^+$	102
6.28. Azul: Distribución de probabilidad de B_m con el corte ópti- mo de BDT. Magenta: Distribución de probabilidad de B_m con el corte óptimo de BDT & PID del decaimiento $B^+ \rightarrow$ $\pi^+ \pi^+ \pi^-$	103

6.29. Azul: Distribución de probabilidad de B_m con el corte óptimo de BDT. Magenta: Distribución de probabilidad de B_m con el corte óptimo de BDT & PID del decaimiento $B^+ \rightarrow K^+ K^- \pi^+$	104
A.1. Verificación del overtraining	111
B.1. Curvas ROC de variables ProbNN para el canal $B^+ \rightarrow \pi^+ \pi^+ \pi^-$	113
B.2. Curvas ROC de variables ProbNN para el canal $B^+ \rightarrow K^+ K^- \pi^+$	114

Índice de Tablas

4.1. Estadísticas de señales de Monte Carlo (MC)	47
4.2. Algunos criterios de selección de stripping 20	51
6.1. Algunos de los posibles cortes para la salida de BDT con sus atributos para el canal $B^+ \rightarrow \pi^+ \pi^+ \pi^-$	91
6.2. Algunos de los posibles cortes para la salida de BDT con sus atributos para el canal $B \rightarrow K^+ K^- \pi^+$	94
6.3. Criterio de selección PID	102
A.1. Variables discriminantes utilizadas en el entrenamiento de MVA, ordenadas de-crecientemente según su poder de separación para el decaimiento $B \rightarrow \pi^+ \pi^+ \pi^-$	108
A.2. Variables discriminantes utilizadas en el entrenamiento de MVA, ordenadas de-crecientemente según su poder de separación para el decaimiento $B \rightarrow K^+ K^- \pi^+$	109

Capítulo 1

Introducción

El experimento LHCb ubicado en el Centro Europeo para la Investigación Nuclear (CERN, por sus siglas en inglés) tiene como objetivo el estudio del fenómeno de violación de la simetría de Carga - Paridad o violación CP y de decaimientos raros en hadrones con contenidos de quarks b y c . La violación CP es uno de los ingredientes claves para entender la diferencia de materia y antimateria en el universo. El CERN es uno de los mayores y más prestigiosos centros de investigación en el mundo ubicado en la frontera entre Francia y Suiza. Es de resaltar que el Gran Colisionador de Hadrones es el acelerador de partículas más grande y con mayor energía en el mundo.

Grandes cantidades de partículas son producidos durante las colisiones protones-protones en el Gran Colisionador de Hadrones (LHC), dichas partículas dejan registrado su paso a través del detector. Un tipo particular de decaimiento, de mucho interés, es de los mesones B en piones y kaones, el cual no posee en su estructura el quark encanto, como ser $B^+ \rightarrow \pi^+\pi^+\pi^-$, $B^+ \rightarrow K^+K^-\pi^+$, $B^+ \rightarrow K^+\pi^+\pi^-$ y $B^+ \rightarrow K^+K^+K^-$ ¹ [35]. Estos decaimientos son altamente sensibles a violación CP y por tanto ofrecen un laboratorio para el estudio de nueva física. El análisis de datos conteniendo estos decaimientos, cantidades de la orden de TB, es de alta complejidad. Uno de los pilares fundamentales en el éxito del estudio de estos procesos es la correcta identificación de las partículas en el estado final

¹La carga conjugada se sobreentenderá en todo el texto

(π y K) y entonces de la correcta identificación de los diferentes canales de decaimiento. Las técnicas de Machine Learning son también aplicadas en la reducción de ruido blanco o combinatorial. Este ruido y otros contaminan la región de señal quedando abajo o alrededor de ella, lo que implica que no sea una distribución Gaussiana como es esperado. La identificación y aislamiento de ese ruido no es fácil, porque han perdido una partícula o se les ha asignado una identidad que no es la de ellas. Esto se convierte en la parte esencial de este estudio y es una pieza clave para la obtención de una muestra limpia de señal.

Este trabajo describe un análisis estadístico, se utilizaron datos recopilados por la Colaboración LHCb en el LHC durante el Run II, que tuvo lugar del año 2015 - 2018. De este dataset se usó únicamente los datos 2015 y 2016. El estudio tiene como objetivo la aplicación de algoritmos de machine learning y la identificación de partículas, para la reducción de ruido en los canales $B^+ \rightarrow h^+h'^+h''^-$ ². Serán utilizado como primer paso los métodos de análisis multivariante de variables que sean buenas discriminantes entre señal y ruido. Se realizará un estudio de eficiencia de cortes en las variables de identificación, también son construidas figuras mérito, donde se maximiza la significancia estadística.

A lo largo de esta disertación el orden utilizado para el canal $B^+ \rightarrow K^+\pi^+K^-$, según la convención descrita en $B^+ \rightarrow h^+h'^+h''^-$ sera $B^+ \rightarrow K^+K^-\pi^+$ entendiéndose que se refiere al mismo decaimiento y solo el orden de las partículas hijas cambia.

Esta disertación esta organizada de la siguiente manera: En el Capítulo 2 se presenta la motivación detrás del estudio y una breve descripción de los fundamentos teóricos relacionados con nuestro análisis. En el Capítulo 3 es descrito el experimento LHCb. En el Capítulo 4 se discuten los pasos del proceso de selección de datos en los decaimientos $B^+ \rightarrow h^+h'^+h''^-$. En el Capítulo 5 se presenta los algoritmos de Machine Learning basados en análisis multivariante. Finalmente los resultados están dados en el Capítulo 6 y las conclusiones en el Capítulo 7. Posteriormente hay dos apéndices, el apéndice (A) de análisis multivariante, donde es descrito las variables

² $B^+ \rightarrow h^+h'^+h''^-$, donde $hh'h''$ representa kaones y piones

CAPÍTULO 1. INTRODUCCIÓN

discriminantes utilizadas en el estudio (A.1) y un análisis del overtraining de BDT (A.2), por ultimo en el apéndice B se muestran las curvas ROC de las variables de identificación de partículas.

Capítulo 2

Fundamentos Teóricos

Este capítulo proporciona una introducción teórica para los temas que son relevantes en esta disertación. En la sección 2.1 se brinda la motivación para este análisis. Seguido, en la sección 2.2, se da una breve introducción del Modelo Estándar de la Física de Partículas terminando en la sección 2.3 con conceptos de estadística.

2.1. Motivación

El Modelo Estándar es el modelo teórico de la Física de Partículas que describe con gran éxito las partículas fundamentales conocidas y las interacciones entre ellas. Dentro de la clasificación de las partículas subatómicas existen los llamados mesones los cuales están constituidos por un quark y un antiquark. El Experimento LHCb (Large Hadron Collider Beauty) fue diseñado para el estudio de los mesones B estos son partículas que contienen el quark- b uno de los quarks mas pesados conocidos y son altamente inestables con tiempos de vida muy cortos. Este tipo de mesones se producen muy raramente y se necesitan grandes aceleradores de partículas para poder estudiarlos; en el detector LHCb se pueden reconstruir con gran precisión al acelerar las partículas con velocidades cercanas a la de la luz y hacerlas colisionar. Debido a estas altas energías de aceleración los mesones son producidos.

Grandes cantidades de partículas son producidas durante las colisiones protón - protón en el LHC. Un tipo particular de decaimiento de mucho interés son los mesones B que no poseen en su contenido de quarks el quark encanto.

Como ser los cuatro canales anteriormente mostrados, en la Figura (2.1) $B^+ \rightarrow K^+ \pi^+ \pi^-$ en el lado superior derecho, $B^+ \rightarrow K^+ K^+ K^-$ en el lado superior izquierdo, $B^+ \rightarrow \pi^+ \pi^+ \pi^-$ en el lado inferior derecho, $B \rightarrow K^+ K^- \pi^+$ en el lado inferior izquierdo. Estos canales fueron analizados con el dataset denominado Run 1 el cual consiste en los datos colectados en 2011 y 2012 y con su estadística fue posible encontrar una alta violación CP en los cuatro canales y mas aún violación CP por regiones del espacio de fase de esos canales [35].

Estos gráficos denotan la representación visual del espacio de fase, esta dividido en bins de dos dimensiones. Es un bin adaptativo que se fue escogiendo de modo que quedara el mismo numero de eventos en cada bin y luego se encontró lo que sería la asimetría CP con correcciones de eficiencia y su extracción de ruido.

Del análisis del Run 1 (Phys.Rev.D90(2014), no.11, 112004 [35])

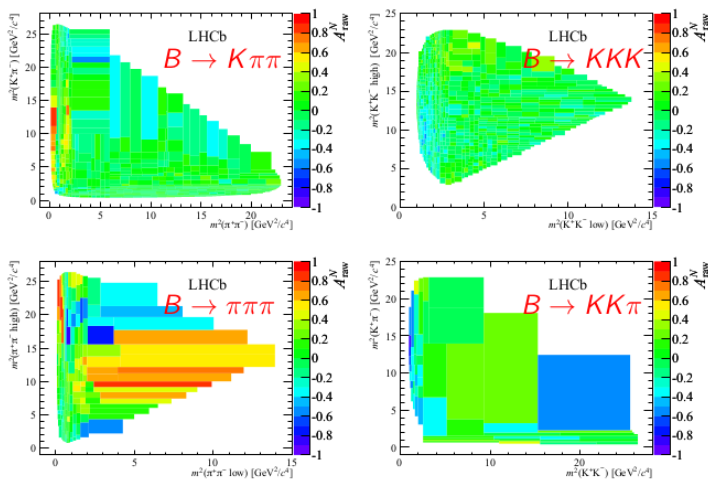


Figura 2.1: Grandes asimetrías en regiones del espacio de fase

Si miramos la Figura 2.1 no esta del mismo color, es decir, si existie-

se el mismo número de B^+ y B^- esos gráficos serían de color verde. La distribución de colores indica la existencia de asimetría, esto debido a la violación CP.

Con la gran y mayor estadística colectada por el detector de 2015 - 2018 (Run 2), es posible volver a analizar estos canales con una cantidad de datos sin precedentes representando el doble de la cantidad que se tenían en el Run 1. Con estos nuevos datos es posible explorar nuevas regiones y mejorar las técnicas utilizadas.

Una parte fundamental en todo análisis con procesos de decaimiento es la correcta clasificación de los datos, porque no solo se tiene eventos de señal también se tiene ruido, mas aún, puede existir distintos tipos de ruido como ser: *combinatorial*, *mis-identificación de partículas* que se confunden con los procesos que se pretenden analizar y *decaimientos parcialmente reconstruidos*. Todo este tipo de ruido afecta la señal y es necesario identificarlos y aislarlos. Por eso la gran importancia de la correcta clasificación de datos particularmente la reducción al máximo de los eventos de ruido pero con el compromiso de mantener nuestra señal.

2.2. Modelo Estándar

“¿Cómo nació el universo y cómo se formó la materia que compone las estrellas, los planetas? ¿De qué está hecha la materia? ¿Que la mantiene unida? ¿Cómo se origina la masa de las partículas elementales, sin la cual no se podrían formar átomos? ¿Y que la materia que nos compone tiene una gemela casi idéntica, la antimateria, que desapareció misteriosamente tras el Big Bang?” [19]

Para investigar las respuestas a estas y otras preguntas se encarga la Física de Partículas. Es interesante, que cuanto mas pequeño sea el objeto a estudiar (partículas elementales), se necesitan instrumentos de medición y de análisis mas grandes y complejos, un ejemplo de ello es El Gran Colisionador de Hadrones.

Por lo tanto, la Física de Partículas tiene como objetivo el estudio de los elementos más fundamentales de la naturaleza que no tienen una estructura interna de la orden de 10^{-15} , 10^{-18} metros, tamaños mucho menores de lo

que es un átomo.

Es fascinante como a lo largo de toda la historia se han realizado muchos avances científicos producto del trabajo de grandes físicos y matemáticos tanto teóricamente como experimentalmente que han dado un panorama mas claro sobre como esta estructurada la materia y como funciona nuestro mundo. El Modelo Estándar de la Física de Partículas es en la opinión de muchos físicos la mejor teoría sobre la naturaleza de la materia y uno de los mas grandes logros del siglo XX. Por ejemplo según Gordon Kane [29], un físico teórico de la Universidad de Michigan:

“El Modelo Estándar es, en la historia, la más sofisticada teoría matemática sobre la naturaleza. A pesar de la palabra “modelo” en su nombre, el Modelo Estándar es una teoría comprensiva que identifica las partículas básicas y especifica cómo interactúan. Todo lo que pasa en nuestro mundo (excepto los efectos de la gravedad) es resultado de las partículas del Modelo Estándar interactuando de acuerdo con sus reglas y ecuaciones (p.58).”

El Modelo Estándar es una de las teorías mas exitosas de la humanidad y que se sigue probando exitosamente. Sin embargo, no es teoría totalmente completa, debido a que la interacción gravitacional esta notablemente ausente del modelo. Esto deja algunos fenómenos inexplicables y poco entendidos. Han surgido varias especulaciones teóricas que intentan explicar estas deficiencias como ser la denominada Física mas allá del Modelo Estándar: como ser la supersimetría, la teoría de las supercuerdas entre otras. Aunque estas teorías no han sido probadas experimentalmente han tenido gran auge en las actuales investigaciones físicas tanto teóricas como experimentales.

En el LHC y LHCb, sus resultados siguen siendo compatibles con la teoría, pero el objetivo de estos detectores en estas eras de energía es ir buscando nueva física. Entonces concluyendo, la Física de Partículas es la rama de la física que estudia los componentes elementales de la materia y las interacciones entre ellas.

En la Figura 2.2 se observa la clasificación de las partículas elementales de la naturaleza. El Modelo Estándar se basa en tres familias de partículas principales en las que tanto los quarks como los leptones forman la materia existente y además se dividen en tres generaciones de la materia (fermio-

nes), descritas a continuación.

Familia de quarks, I: up, down; II: charm, strange y III: top, bottom. Familia de leptones, I: electrón, neutrino de electrón. II: muon, neutrino de muon. III: tau, neutrino de tau. Mientras que los bosones actúan como partículas virtuales y describen las interacciones entre las partículas mediante campos de fuerzas. Las partículas denominadas Bosones de las fuerzas fundamentales, el **Fotón** de la Interacción electromagnética. Los W^+ , W^- , Z^0 de la Interacción débil que es la responsable de los decaimientos o desintegraciones. Los **Gluones**, es la partícula mediadora de la interacción fuerte, es la que permite mantener los quarks unidos para formar partículas.

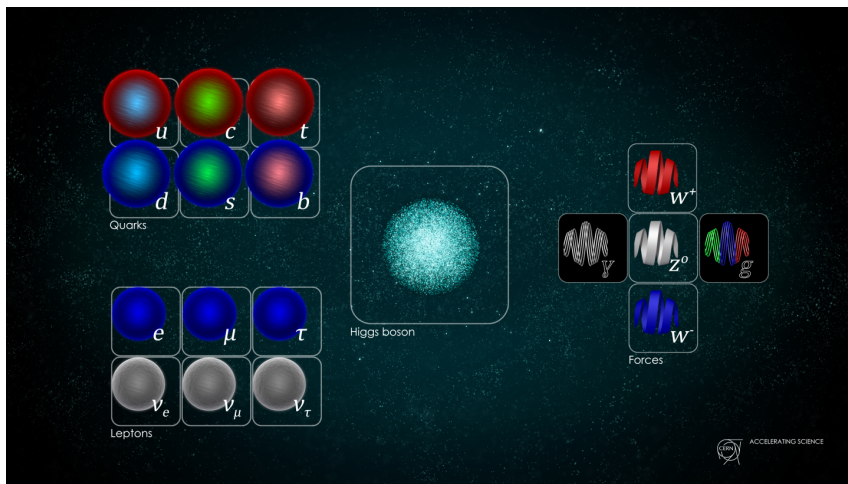


Figura 2.2: Partículas elementales e interacciones fundamentales en la naturaleza

Se incluye también la partícula llamada el Bosón de Higgs con la que se entiende que las partículas al interactuar con el campo de Higgs adquieren su masa. Para mayor información del Modelo Estándar ver [25]

Los canales $B^+ \rightarrow h^+ h' + h''^-$ tienen un su contenido de quark a, 'b' y son caracterizados por transiciones $b \rightarrow u$ & $b \rightarrow d$

Los quarks se agrupan para formar partículas llamadas hadrones. En la naturaleza se han observado dos tipos de hadrones, los cuales son bariones y mesones. Los bariones están compuestos por tres quarks o tres anti-quarks, ejemplos de ellos son: el protón (uud) y el neutrón (udd). Los mesones están compuestos por un quark y un anti-quark, ejemplos de ellos son: los piones (π^\pm) contienen solo quarks up y down y kaones (K^\pm) contienen un quark strange y un quark up. Los decaimientos en estudio de tres cuerpos $B^+ \rightarrow \pi^+\pi^+\pi^-$ y $B^+ \rightarrow K^+K^-\pi^+$ son mesones.

El mesón B^+ también llamado de beauty, tiene en su contenido un antiquark beauty y un quark up ($u\bar{b}$), el tiempo de vida del mesón B es de $1,638 \times 10^{-12}$ s y la masa proyectada es de $5279,29 \pm 0,15$ MeV [26]. El nombre “sin encanto” se refiere a la ausencia del quark encanto en el estado final del decaimiento. Las desintegraciones sin encanto son un proceso poco probable, ya que implican la transformación de un quark beauty en un quark up y predominantemente el quark beauty se desintegra en un quark charm.

2.3. Conceptos Estadísticos

Las características físicas relacionadas con las colisiones protón-protón (pp) hasta el decaimiento de interés son guardadas en ‘eventos’ como ser, momentos de partículas, numero de muones, parámetros de impacto, distancia de vuelo, variables de identificación de partículas ,..., etc. A partir de ello se necesita estimar parámetros, cuantificar incertidumbres en las estimaciones, evaluar que tan bien estima una teoría determinada con los datos observados y para todo ello nos ayuda la estadística. En este trabajo se dará por hecho que se conocen las nociones básicas de probabilidad (que es muy importante en la solución de problemas de machine learning) y comenzaremos a definir algunas distribuciones de probabilidad mas utilizadas en Física de Partículas.

2.3.1. Distribuciones de probabilidad

En esta sección, se presenta una breve descripción de las distribuciones de probabilidad que son utilizadas en la modelización de procesos físicos.

Como ser, distribución Bernoulli, Binomial y Multinomial para variables aleatorias discretas las cuales son fundamentales para problemas de clasificación. Las distribuciones, Uniforme, Exponencial y Gaussiana para variables aleatorias continuas. Estas distribuciones son paramétricas, porque están gobernadas por un pequeño número de parámetros. Por ejemplo, la distribución normal tiene como parámetros a la media y la varianza. Se necesita estimar los valores adecuados para estos parámetros, dado un conjunto de datos observado. En el enfoque frecuentista, son seleccionados valores específicos para los parámetros utilizando por ejemplo la función de verosimilitud, mientras que para un enfoque bayesiano se dan distribuciones prioris sobre los parámetros para luego utilizar el teorema de Bayes y así encontrar la distribución posterior correspondiente dados los datos. Para más detalle consultar [5], [18] y [32].

1. Variables aleatorias binarias

Consideremos una variable aleatoria binaria $X \in \{0, 1\}$. Especialmente útil para problemas de clasificación binarios.

Sea $P(X = 1|\mu) = \mu$ y $P(X = 0|\mu) = 1 - \mu$ donde $0 \leq \mu \leq 1$. Por tanto, la distribución de probabilidad sobre x se puede escribir en la forma

$$\text{Bern}(X|\mu) = \mu^x(1 - \mu)^{1-x} \quad (2.3.1)$$

Para calcular los momentos de la distribución, tenemos que

$$E[X] = \sum_{x=0,1} x \text{Bern}(x|\mu) = \sum_{x=0,1} x \mu^x (1 - \mu)^{1-x} = \mu \quad (2.3.2)$$

$$V(X) = E[X^2] - (E[X])^2 = \mu - \mu^2 = \mu(1 - \mu) \quad (2.3.3)$$

Suponemos que tenemos un conjunto de datos $\mathcal{D} = \{x_1, x_2, \dots, x_N\}$ de los valores observados de X . Construimos la función de verosimilitud en función del parámetro μ , asumiendo que cada una de los valores de la muestra son independientes e idénticamente distribuidas (iid) de $p(X|\mu)$, tal que

$$P(\mathcal{D}|\mu) = \prod_{i=1}^N p(x_i|\mu) = \prod_{i=1}^N \mu^{x_i} (1 - \mu)^{1-x_i} \quad (2.3.4)$$

En el enfoque frecuentista podemos estimar un valor para μ maximizando la función de verosimilitud o de manera similar maximizando el logaritmo de la probabilidad. Así la función logarítmica de la verosimilitud viene dada por:

$$\ln P(\mathcal{D}|\mu) = \sum_{i=1}^N \ln p(x_i|\mu) = \sum_{i=1}^N \{x_i \ln \mu + (1 - x_i) \ln(1 - \mu)\} \quad (2.3.5)$$

Si obtenemos la derivada de $\ln P(\mathcal{D}|\mu)$ con respecto a μ y la igualamos a cero, obtenemos el estimador de máxima verosimilitud

$$\begin{aligned} \frac{d}{d\mu} \ln P(\mathcal{D}|\mu) &= \frac{d}{d\mu} \left[\sum_{i=1}^N x_i \ln \mu + (N - \sum_{i=1}^N x_i) \ln(1 - \mu) \right] \\ &= \frac{\sum_{i=1}^N x_i}{\mu} - \frac{(N - \sum_{i=1}^N x_i)}{1 - \mu} = 0 \\ (1 - \mu) \sum_{i=1}^N x_i - \mu(N - \sum_{i=1}^N x_i) &= 0 \end{aligned} \quad (2.3.6)$$

Así tenemos que,

$$\mu_{ML} = \frac{\sum_{i=1}^N x_i}{N} \quad (2.3.7)$$

La probabilidad de obtener m eventos de N posibles en donde $x = 1$, esta dado por la **distribución binomial**

$$Bin(m|N, \mu) = \binom{N}{m} \mu^m (1 - \mu)^{N-m} \quad (2.3.8)$$

Dado que $m = x_1 + x_2 + \dots + x_N$ y para cada observación la media y la varianza están dadas por 2.3.9 y 2.3.10, tenemos que

$$E[m] = E \left[\sum_{i=1}^N x_i \right] = \sum_{i=1}^N E[x_i] = N\mu \quad (2.3.9)$$

$$V[m] = V \left[\sum_{i=1}^N x_i \right] = \sum_{i=1}^N V[x_i] = N\mu(1 - \mu) \quad (2.3.10)$$

La esperanza matemática para la distribución binomial también puede obtenerse de la siguiente forma, haciendo uso de la identidad $m \binom{N}{m} = N \binom{N-1}{m-1}$:

$$\begin{aligned} E[m] &= \sum_{m=0}^N m \binom{N}{m} \mu^m (1 - \mu)^{N-m} \\ &= \sum_{m=1}^N m \binom{N}{m} \mu^m (1 - \mu)^{N-m} \\ &= \sum_{m=0}^N N \binom{N-1}{m-1} \mu^m (1 - \mu)^{N-m} \\ &= \sum_{y=0}^{N-1} N \binom{N-1}{y} \mu^{y+1} (1 - \mu)^{N-(y+1)} \\ &= N\mu \sum_{y=0}^{N-1} \binom{N-1}{y} \mu^y (1 - \mu)^{N-1-y} \\ &= N\mu \end{aligned} \quad (2.3.11)$$

Esta distribución es sutil para problemas de clasificación de dos clases como por ejemplo separa señal y ruido.

2. Variables aleatorias multinomiales

No siempre nuestras variables van a ser binarias. Tomado de [5]. En muchas ocasiones hay variables que pueden tomar varios valores mutuamente excluyentes. Podemos expresar esta variable como un vector X de dimensión K , en el cual uno de los elementos x_k es igual a 1 y todos los demás elementos son iguales a cero. Por ejemplo de $K = 5$ y $x_2 = 1$, entonces el vector X es:

$$X = (0, 1, 0, 0, 0)^T \quad (2.3.12)$$

tal que, $\sum_{k=1}^K x_k = 1$ y $P(x_k = 1 | \mu_k) = \mu_k$, entonces la distribución de X esta dado por:

$$P(X|\mu) = \prod_{k=1}^K \mu_k^{x_k} \quad (2.3.13)$$

Donde $\mu = (\mu_1, \dots, \mu_K)^T$ y $\sum_k \mu_k = 1$. La distribución 2.3.13 es una generalización de la distribución de Bernoulli. Ahora bien, consideremos el conjunto de datos \mathcal{D} de N observaciones independientes x_1, x_2, \dots, x_N . La función de verosimilitud es la siguiente:

$$P(\mathcal{D}|\mu) = \prod_{n=1}^N \prod_{k=1}^K \mu_k^{x_{nk}} = \prod_{k=1}^K \mu_k^{(\sum_n x_{nk})} = \prod_{k=1}^K \mu_k^{m_k} \quad (2.3.14)$$

Ahora consideremos la distribución conjunta de m_1, \dots, m_K , asi la función de verosimilitud toma la forma

$$Mult(m_1, \dots, m_K | \mu, N) = \binom{N}{m_1, m_2, \dots, m_K} \prod_{k=1}^K \mu_k^{m_k} \quad (2.3.15)$$

La ecuación 2.3.15 es la llamada distribución multinomial. El coeficiente de normalización es el numero de formas de dividir N objetos en K grupos de tamaño m_1, m_2, \dots, m_k y viene dada por:

$$\binom{N}{m_1 m_2, \dots, m_K} = \frac{N!}{m_1! m_2! \dots m_K!} \quad (2.3.16)$$

Se debe tener en cuenta que las variables m_k están sujetas a la restricción:

$$\sum_{k=1}^K m_k = N \quad (2.3.17)$$

3. Distribución Uniforme:

Considere una variable aleatoria continua X , con una función densidad de probabilidad $P(X|a, b)$ definida dentro de un intervalo finito $[a, b]$:

$$P(X|a, b) = \begin{cases} \frac{1}{b-a}, & a \leq X \leq b \\ 0, & \text{en otro caso} \end{cases} \quad (2.3.18)$$

Para esta distribución la media y la varianza son:

$$E[X] = \frac{a+b}{2} \quad (2.3.19)$$

$$V[X] = \frac{(b-a)^2}{12} \quad (2.3.20)$$

Se puede decir que un enfoque sencillo de generación de Monte-Carlo se basaría en una distribución de probabilidad uniforme en el rango de $[0, p]$, donde p es la probabilidad de éxito de la implementación de un algoritmo de aceptación-rechazo, que usaría valores de x generados aleatoriamente en el rango de $[0, 1]$.

4. Distribución Exponencial:

Considere una variable continua X , con una función densidad de probabilidad $P(X; \beta)$ dada por:

$$P(X|\beta) = \begin{cases} \frac{1}{\beta} \exp^{-\frac{x}{\beta}}, & X \geq 0 \\ 0, & \text{en otro caso} \end{cases} \quad (2.3.21)$$

Para esta función de probabilidad la media y la varianza son:

$$E[X] = \beta \quad (2.3.22)$$

$$V[X] = \beta^2 \quad (2.3.23)$$

Se dice que la distribución exponencial no tiene memoria. Una aplicación importante en física de altas energías de esta distribución es la descripción de fenómenos que ocurren de forma independiente a un ritmo constante, como la duración de los decaimientos y su vida útil.

5. Distribución Gaussiana:

La distribución Gaussiana, también conocida como distribución normal, es una de las distribuciones más utilizadas en estadística para variables aleatorias continuas. En el caso de una sola variable X , la distribución gaussiana se puede escribir de la siguiente forma:

$$\mathcal{N}(X|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\} \quad (2.3.24)$$

Donde la media es μ y la varianza es σ^2 . Ahora bien, para un vector X D -dimensional, la distribución gaussiana multivariante es la siguiente:

$$\mathcal{N}(X|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right\} \quad (2.3.25)$$

Donde μ es el vector medio D -dimensional, Σ es la matriz de covarianzas de orden $D \times D$

La distribución Gaussiana supera a las demás distribuciones tanto conceptualmente como en aplicaciones prácticas, esto se debe en gran

parte al Teorema del Limite Central, que asegura que las distribuciones Gaussianas son el límite de los procesos que surgen de múltiples fluctuaciones aleatorias. Si consideramos n variables aleatorias independientes e idénticamente distribuidas $X = (X_1, X_2, \dots, X_n)$ cada una con media y varianza μ y σ^2 , respectivamente, la media de esas variables en el límite tiene una distribución normal estándar.

2.3.2. Pruebas de hipótesis

Las pruebas de hipótesis en el contexto de Física de Partículas son una de las principales herramientas estadísticas para la búsqueda de señales cada vez mas pequeñas entre una gran cantidad de ruido. Una parte crucial en un análisis estadístico es probar si las predicciones de un modelo dado están de acuerdo con los datos observados. Una hipótesis H es una suposición acerca de los datos observados. Denotaremos con X al estadístico de prueba que es función de las mediciones de la muestra (datos observados), X puede representar una sola cantidad o colección de valores. Sea $P(X|H)$ (Si X contiene variables continuas, entonces la función de densidad de probabilidad es denotada por $f(X|H)$) la probabilidad de encontrar datos X bajo el supuesto de la hipótesis H .

En estadística existen dos tipos de hipótesis: H_0 denominada hipótesis nula y H_1 hipótesis alternativa, estas hipótesis en el contexto físico son:

- Selección de eventos: Donde se sabe que los eventos a tratar existen. Por ejemplo: H_0 : el evento es ruido vs H_1 : el evento es señal. Al final los eventos seleccionados se utilizan para estudios posteriores.
- Búsqueda de nueva física:
 H_0 : Todos los eventos son del Modelo Estándar (solo ruido)
 H_1 : Los eventos incluyen un tipo de existencia aún no establecida (señal mas ruido).
 Un ejemplo conocido es el descubrimiento del Bosón de Higgs

En este análisis se ha utilizado la selección de eventos con el propósito de que los eventos seleccionados sean utilizados en posteriores estudios.

Ahora bien, definimos una prueba de H_0 determinando un subconjunto del espacio de datos al cual denominamos región crítica w , son los valores del estadístico de prueba que permiten rechazar la hipótesis nula, tal que la probabilidad de que los datos estén ahí cumple $P(X \in w|H_0) \leq \alpha$. El valor de α se especifica antes de la prueba y usualmente se utiliza valores pequeños, por ejemplo, 1 %, 5 % y 10 %. La región crítica se establece de que halla la mayor probabilidad de encontrar los datos ahí si la H_1 es verdadera, esto se ilustra en la Figura 2.3

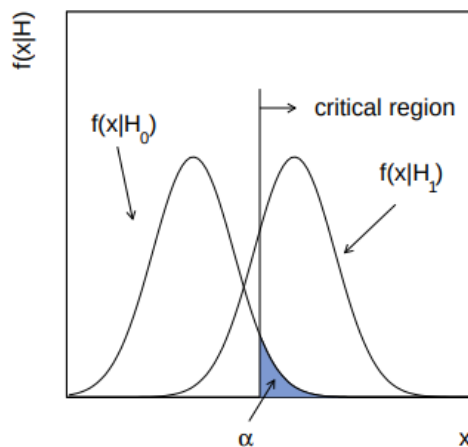


Figura 2.3: Región crítica de una prueba estadística (Tomado de [18])

Una prueba de hipótesis nos lleva a cometer errores ya que se toma una decisión basada en los datos observados. Rechazar H_0 cuando es verdadera se denomina error tipo I y la probabilidad de que este ocurra es $P(X \in w|H_0) \leq \alpha$. Cuando no rechazamos H_0 cuando H_0 es falsa se tiene el error tipo II y la probabilidad de que este ocurra es $P(X \in S - w|H_1) = \beta$. Definimos también la potencia de la prueba con respecto a H_1 como *potencia* = $1 - \beta$ que es la probabilidad de rechazar la hipótesis nula si la alternativa H_1 es verdadera. En un enfoque frecuentista, rechazamos H_0 si los datos están en la región crítica, es decir, si el valor p es menor o igual a α .

Ahora bien, para determinar el límite de decisión de la región crítica en un conjunto de n variables, se define la función $y(x_1, \dots, x_n)$ como una

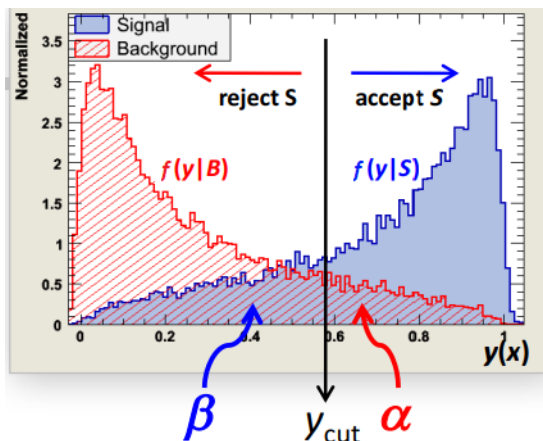


Figura 2.4: Distribuciones de probabilidad (pdf) del estadístico de prueba escalar $y(x_1, \dots, x_n)$ bajo las hipótesis de señal y ruido.

estadística de prueba escalar. Así, las pdf de $y(x_1, \dots, x_n)$ bajo las hipótesis H_0 : el evento es ruido vs H_1 : el evento es señal; son respectivamente $f(y|b)$ y $f(y|s)$. Así, el límite de decisión es ahora un solo corte en la variable y , mismo que se ilustra en la Figura 2.4.

Retomando los tipos de errores, la *probabilidad de mis-identificación de ruido* corresponde a error tipo I (α), es decir, se da cuando un ruido es identificado como señal y el error tipo II (β) es igual a uno menos la eficiencia de la señal, es decir, la probabilidad de identificar incorrectamente una señal como un ruido. La fracción esperada de partículas de señal seleccionadas generalmente es denominada *eficiencia de selección de señal o eficiencia de la señal* y la fracción esperada de partículas de ruido seleccionadas es denominada *probabilidad de identificación errónea (mis-identificación)*.

Si variamos el valor del corte y_{cut} de la Figura 2.4, se pueden determinar diferentes valores de eficiencia de señal ($1 - \beta$) y de la probabilidad de mis-identificación ($1 - \alpha$). En la Figura 2.5 obtenida al variar el corte y_{cut} se muestra una curva que representa la eficiencia de la señal frente a la probabilidad de mis-identificación, este gráfico es llamado característica

operativa del receptor o curva ROC.

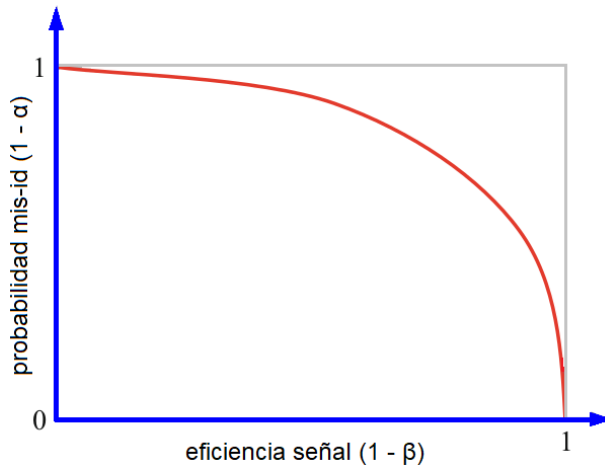


Figura 2.5: Curva ROC: Eficiencia de la señal frente a la probabilidad de mis-identificación de ruido.

Se busca que la selección del corte óptimo tenga una baja probabilidad de mis-identificación que involucre una gran eficiencia de la señal. Esta claro que la probabilidad de mis-identificación no puede ser cero, es decir, el rechazo de ruido no puede ser perfecto, si las distribuciones de probabilidad $f(X|H_0)$ y $f(X|H_1)$ se superponen como se muestra en la Figura 2.4.

2.3.3. El Lema de Neyman-Pearson

Como se mencionó en la sección 2.3.2 el rendimiento de un criterio de selección es óptimo cuando logra la menor cantidad de mis-identificación para una gran eficiencia de la señal. El lema de Neyman-Pearson nos dice que el estadístico de prueba óptimo, viene dado por el cociente de las funciones de verosimilitud $L(x|H_0)$ y $L(x|H_1)$, evaluada en la muestras de datos observados x bajo las hipótesis H_0 y H_1 .

$$\lambda(x) = \frac{L(x|H_1)}{L(x|H_0)} \tag{2.3.26}$$

Esta prueba es óptima, cuando se fija la probabilidad de mis-identificación de ruido α , la selección que corresponde a la mayor eficiencia de la señal $1 - \beta$ es:

$$\lambda(x) = \frac{L(x|H_1)}{L(x|H_0)} \geq k_\alpha \quad (2.3.27)$$

Al variar el valor del ‘corte’ k_α , se puede lograr el valor requerido de α .

El lema plantea que se requiere conocer las distribuciones de probabilidad (PDF), en muchos casos no es fácil determinar las PDF, sin embargo, existen métodos numéricos y algoritmos que brindan resultados cercanos a la selección óptima dada por el lema de Neyman-Pearson. Por ejemplo, algunos algoritmos son de machine learning, como ser, las redes neuronales artificiales y árboles de decisión.

2.3.4. Razón de Verosimilitud Projectiva Discriminante

Sean las variables X_1, \dots, X_n independientes, la función de verosimilitud se puede factorizar en el producto de las distribuciones de probabilidad marginales:

$$\lambda(X_1, \dots, X_n) = \frac{L(X_1, \dots, X_n|H_1)}{L(X_1, \dots, X_n|H_0)} = \frac{\prod_{i=1}^n f_i(x_i|H_1)}{\prod_{i=1}^n f_i(x_i|H_0)} \quad (2.3.28)$$

Según el lema Neyman-Pearson, con ello se logra un desempeño óptimo. Aunque las variables no sean independientes la ecuación 2.3.28, se usa como discriminante utilizando las PDF marginales f_i para cada variable.

Citando a Luca Lista [32] nos dice, en el caso de que las PDF no se puedan factorizar exactamente es decir que no sean independientes, de todos modos, la estadística de prueba definida en la ecuación 2.3.28 diferirá de la razón de verosimilitud exacta de la ecuación 2.3.26 y esto conducirá a rendimientos subóptimos de los dados por el lema de Neyman Pearson. Sin embargo, Luca Lista menciona que en algunos casos, la simplicidad de este método puede justificar su aplicación a pesar de los rendimientos

subóptimos. Las PDF marginales f_i pueden obtenerse utilizando muestras de simulación de Monte Carlo con una gran cantidad de datos de entrada, permitiendo la construcción de histogramas que corresponden a las distribuciones de las variables x_i .

También en su libro Luca Lista menciona “*que algunas aplicaciones numéricas implementan esta técnica de razón de verosimilitud factorizada después de aplicar la transformación de variable adecuada para reducir o eliminar la correlación de las variables. Esto mitiga la disminución del rendimiento en comparación con el límite de Neyman-Pearson, pero no necesariamente permite alcanzar los rendimientos óptimos, porque las variables no correlacionadas no son necesariamente independientes*”.

Capítulo 3

Experimento LHCb

El experimento LHCb es uno de los 4 grandes experimentos del LHC (Gran Colisionador de hadrones). En la sección 3.1 son presentadas generalidades del LHC seguidamente en la sección 3.2 detalla uno de los aspectos importantes acerca de los datos del LHC que es la llamada luminosidad. En la sección 3.3 es descrito el detector LHCb y en la sección 3.4 se especifica los subdetectores del LHCb. En la sección 3.5 contiene información del sistema de Trigger y por ultimo en la sección 3.6 se presentan los sistemas de software del experimento LHCb de una forma breve.

3.1. El Gran Colisionador de Hadrones

El CERN es un laboratorio de Física de Partículas siendo uno de los más grandes y poderosos del mundo donde se hace ciencia de frontera. Esta situado en la frontera entre Francia y Suiza. En el laboratorio están dispuestas una serie de maquinas aceleradoras, siendo la más poderosa de ellos el LHC. El LHC es el acelerador de partículas más potente que hasta la fecha fue diseñado para hacer colisionar haces de protones cada $25 \mu s$, con una energía de 14 TeV (Tern Electrón Voltio) en el referencial del centro de masa.

Con el objetivo de lograr estas energías, el LHC necesita dimensiones físicas considerables. Consta de una serie de estructuras aceleradoras para

aumentar la energía de los haces de protones. Los protones se hacen circular en anillos de 27 km como se ve en la Figura 3.1. El acelerador se encuentra bajo tierra, en un túnel ubicado que varía de 50 a 150 metros. Esta ubicación subterránea del acelerador es estratégicamente necesaria porque protege los datos de la contaminación, es decir, los protege de ruido proveniente de los rayos cósmicos y la radiación, así mismo, también protege a las personas de cualquier mal funcionamiento del acelerador.

Comenzó su funcionamiento por primera vez el 10 de septiembre de 2008 y sigue siendo la última incorporación al complejo de aceleradores del CERN. [7]



Figura 3.1: CERN

El acelerador tiene dos tubos internos que contienen los haces de protones acelerados del LHC. Estos rayos orbitan en direcciones opuestas dentro de los tubos, y son guiados en su trayectoria por imanes que se encuentran en toda la extensión de acelerador.

Los haces tienen paquetes con 100,000 millones de protones. Cuando se aceleran los protones en el túnel de LHC siguen circulando hasta que se les hace llegar a la energía de colisión 14 TeV son nubes de protones y cada nube de lleva de la orden de 100,000 millones de protones y se hacen

CAPÍTULO 3. EXPERIMENTO LHCb

colisionar muchas veces y estratégicamente se hacen colisionar en los puntos donde se encuentran los detectores de modo que ellos puedan registrar toda esa riqueza de partículas creadas en ese momento.

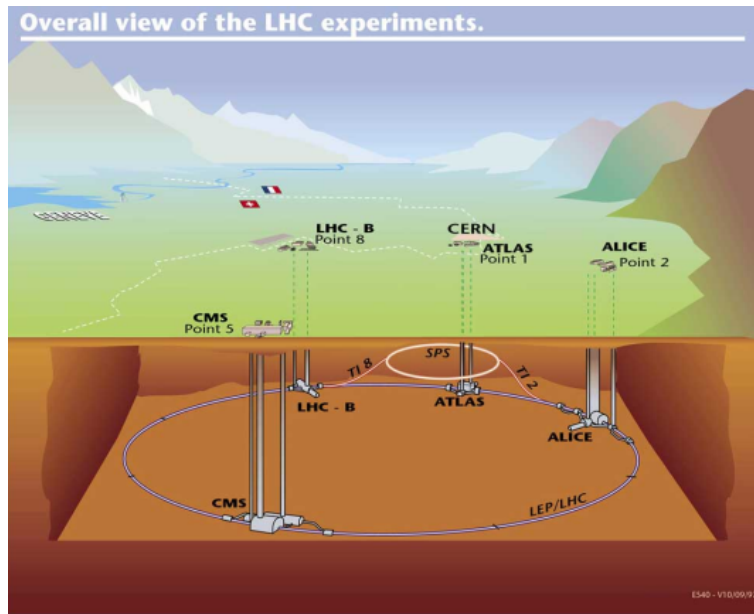


Figura 3.2: Vista esquemática del complejo LHC, que muestra los diferentes puntos de interacción de sus cuatro experimentos principales

Los haces en órbita que se encuentran en los tubos del LC se mantienen separados e interactúan en cuatro puntos a lo largo de la circunferencia del acelerador, donde se encuentran los cuatro detectores más grandes del CERN (ver Figura 3.2). Cuando se logra alcanzar la energía deseada y los rayos chocan, los detectores de estos experimentos, que funcionan simultáneamente, inician su registro de datos. Estos detectores se utilizan en los cuatro experimentos principales del LHC: ALICE, ATLAS, CMS y LHCb. Se presenta una imagen de cada uno de ellos en la Figura 3.3

Explicación breve de cada uno de los experimentos:

ALICE (A Large Ion Collider Experiment): Como su nombre lo indica el detector de este experimento fue diseñado para el estudio de los iones pesados enfocados en la Cromodinámica Cuántica (QCD). El plasma

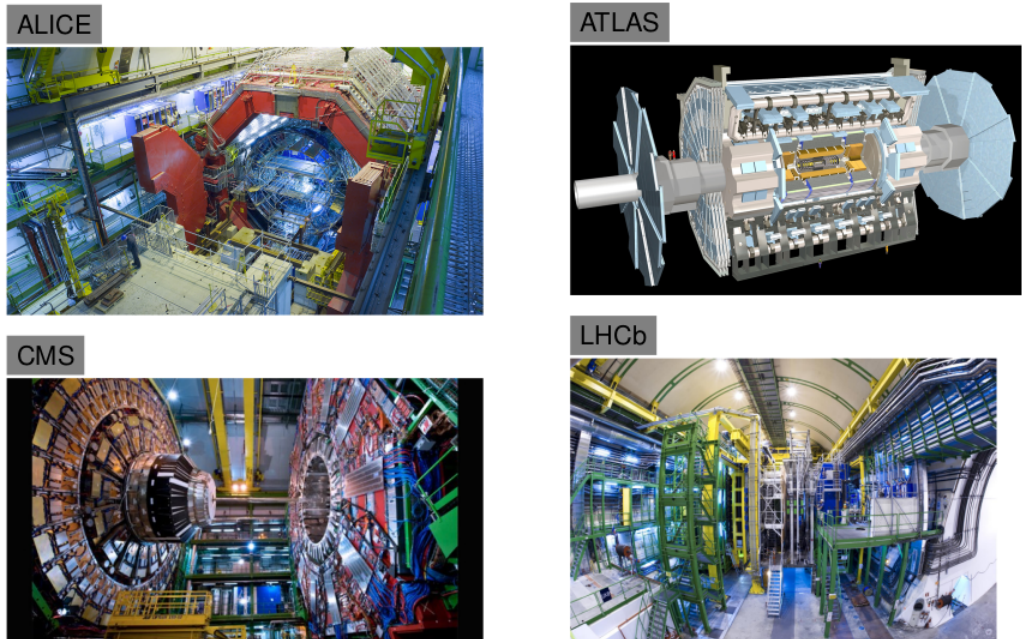


Figura 3.3: Principales experimentos del LHC

quark-gluon se puede alcanzar mediante la interacción fuerte de partículas en condiciones de energía extrema en colisiones Pb-Pb.

CMS (Compact Muon Solenoid) y ATLAS (A Toroidal Lhc ApparatuS): Son los detectores que tienen un propósito general en el LHC el cual es la búsqueda directa de nueva física, dentro de sus objetivos está descubrir nuevas partículas provenientes de las colisiones protón-protón, uno de los más importantes de estos objetivos se logró en julio de 2012 con el descubrimiento del bosón de Higgs, que es una pieza fundamental para completar el modelo estándar. Dentro de sus múltiples objetivos están estudiar las propiedades de Higgs y otros bosones similares a Higgs y buscar indicios de la supersimetría.

LHCb (Large Hadron Collider beauty) [2] y [22]. El experimen-

to LHCb tiene como objetivo el estudio del fenómeno de violación de la simetría de carga - paridad o violación CP y de decaimientos raros en hadrones con contenidos de quarks b y c. La Violación CP es uno de los ingredientes claves para entender la diferencia de materia y antimateria en el universo.

3.2. Datos colectados en el LHC

Uno de los parámetros mas importantes acerca de los datos del LHC es la llamada luminosidad, la cual nos da el resultado de el número de colisiones producidas por el acelerador. Esto se logra encontrar por las luminosidades instantáneas e integradas.

La luminosidad instantánea de un colisionador mide el número de colisiones por unidad de área y tiempo (ver figura 3.4), es un parámetro clave para medir la eficacia del acelerador. Se puede decir que entre mas colisiones hayan en los puntos de interacción hay una mayor probabilidad de que ocurran eventos raros para los cuales fue diseñado el acelerador.

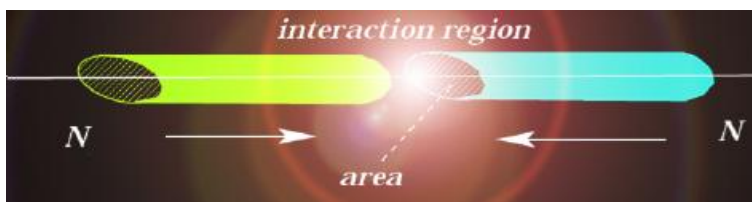


Figura 3.4: La luminosidad nos da el número de colisiones posibles por unidad de área y tiempo.

Se estimó que el LHC funcionaria con una luminosidad instantánea de $\mathcal{L} \approx 10^{34} \text{cm}^{-2} \text{s}^{-1}$. Así, en los puntos de interacción donde se alcance este valor de \mathcal{L} se pueden producir 1034 colisiones por segundo y por cm^2 . Esta estimación es 100 veces mayor que el valor en el acelerador Tevatron en Chicago (USA) (ver [39]), podemos decir que serán mucho más frecuentes los evento raros en el LHC.

También tenemos la luminosidad integrada que es la integral de la luminosidad instantánea sobre el tiempo de funcionamiento del colisionador

$(\mathcal{L}.\Delta t)$ representa el número total de colisiones. En referencia a las colisiones esa probabilidad se conoce como sección transversal (σ)

La siguiente expresión nos da el número de eventos esperados en un determinado tiempo, $N^\circ \text{ eventos} = \text{Sección eficaz} \times \text{Luminosidad integrada}$

$$N^\circ \text{ eventos} = \sigma.\mathcal{L}.\Delta t \quad (3.2.1)$$

Podemos decir, que en el LHC se suman dos parámetros: mayor sección eficaz y mayor luminosidad integrada. En la Figura 3.5 se visualiza la evolución de la luminosidad integrada del experimento LHCb durante 2010 a 2018, involucrando el Run 1 y Run 2.

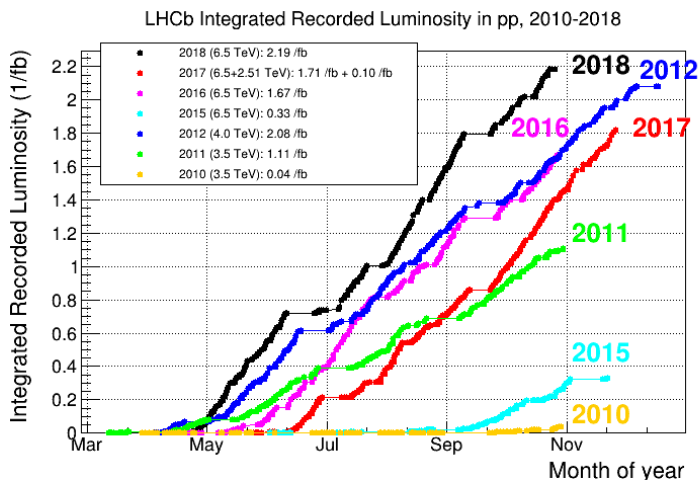


Figura 3.5: Luminosidad integrada producida por el LHC 2010 - 2018

En esta disertación se utilizaron los datos del año 2015 y 2016, podemos observar que la cantidad de luminosidad integrada del 2016 aproximadamente es el doble en comparación al 2015. Para más información sobre la luminosidad del LHCb [20] y [39].

3.3. El Detector LHCb

El detector LHCb es un espectrómetro de un solo brazo especializado en identificar partículas compuestas por el quark b y c , estas partículas

salen disparadas hacia adelante y hacia atrás. En general, podríamos tener el LHCb para un lado y una copia de espejo para el otro lado pero es muy costoso entonces solo se construyó para detectar partículas de un solo lado. El detector cubre una región angular de 10 mrad a 300 en el plano horizontal y de 10 mrad a 250 mrad en el plano vertical como se observa en la ilustración 3.6

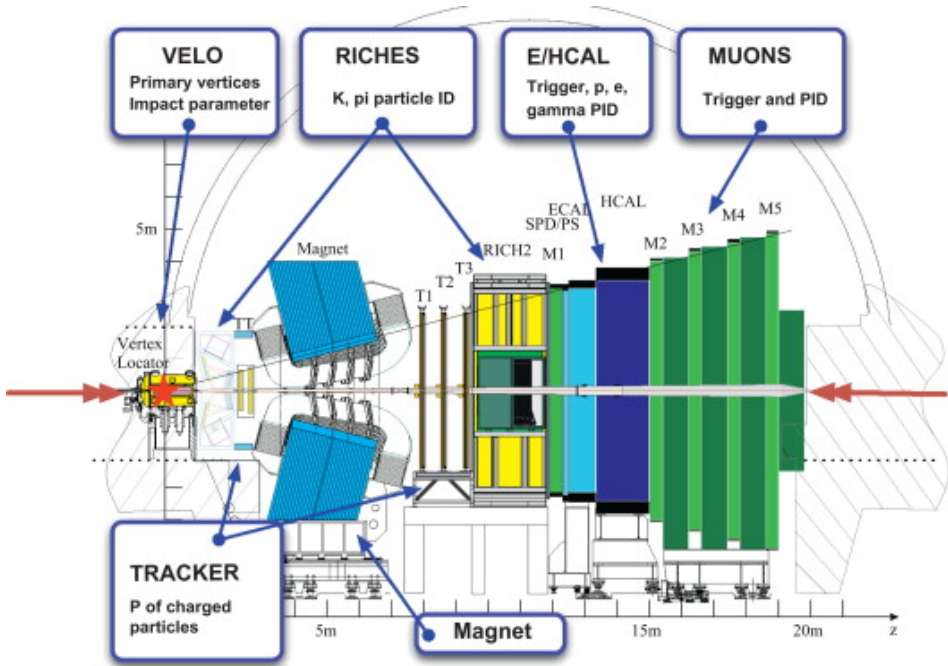


Figura 3.6: Vista lateral del detector LHCb y sus subdetectores

En altas energías el diseño geométrico del detector LHCb se debe a la característica del par $b\bar{b}$ de producirse únicamente en ángulos pequeños, que se ilustra en la Figura 3.7

Cuando ocurre la colisión las partículas salen indistintamente hacia todos lados. Solo aquellos que pasan por el detector o dentro de la aceptación del detector es lo que nosotros conseguimos ver, son exactamente los datos que conseguimos analizar.

El detector LHCb puede describirse como una sucesión de subdetectores

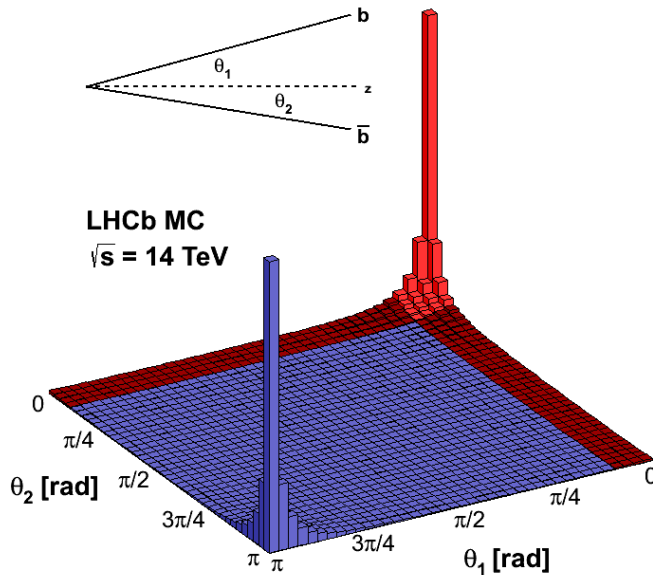


Figura 3.7: En colisiones de alta energía, simulación de Monte Carlo que muestra la producción de pares $b\bar{b}$ en ángulos pequeños.

creados con distintos propósitos, estas son las encargadas de la medición de las trazas de partículas generadas en la colisión protón - protón. En las siguientes secciones veremos con más detalle los subdetectores del LHCb, basándonos en [2] y [22] que contienen una explicación muy detallado del detector LHCb.

3.4. Subdetectores del LHCb

El primer subdetector se localiza en el punto de la interacción y los demás se ubican posteriormente. Estos subdetectores deben ser capaces no solo de medir las trayectorias, los momentos de las partículas y la identifi-

cación de las partículas en sus estados finales, sino también de separar los vértices primarios y secundarios de las desintegraciones, donde el vértice primario es el punto donde se crean las partículas y el vértice secundario es el punto donde decaen.

3.4.1. Vertex Locator (VELO)

El VELO tiene la función de medir con precisión las coordenadas de las pistas que dejan las partículas producidas en el vértice primario ¹ para lograr identificar los vértices secundarios desplazados. La distancia entre los vértices primarios y secundarios es una característica fundamental que permite distinguir los hadrones de belleza y encanto.

El VELO se compone de 42 módulos de detectores de silicio (separados en dos mitades) como se muestra en la Figura 3.8 estos son colocados a una distancia de tan solo 5 mm del punto de interacción del haz de protones.

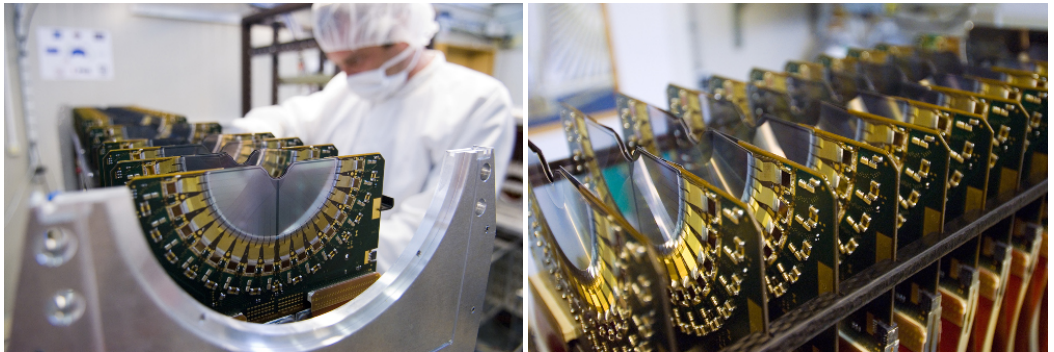


Figura 3.8: Imágenes que muestran una vista exterior del módulo VELO en el LHCb

El sistema VELO puede utilizar la geometría cilíndrica (coordenadas $r\varphi$) con el objetivo de reconstruir las pistas y vértices de una forma más efectiva que los esquemas rectilíneos simples. Cada módulo está diseñado para proporcionar toda la información 3D necesaria para estas reconstrucciones, entonces posee dos sensores distintos: un sensor \mathcal{R} - y φ - como se ilustra en la Figura 3.9

¹Punto donde ocurre la interacción del haz del LHC

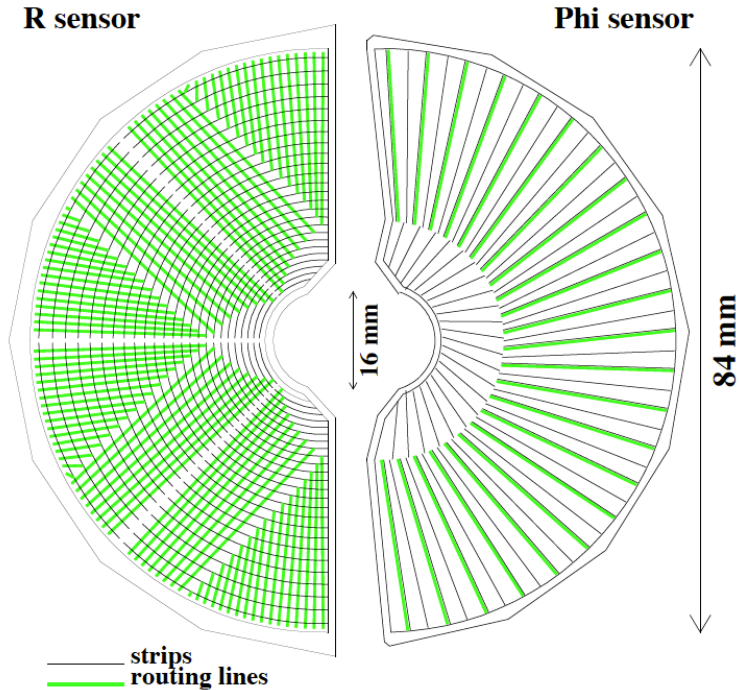


Figura 3.9: Representación del sensor \mathcal{R} - y del sensor φ -.

El sensor φ - proporciona información sobre las coordenadas azimutales del haz y el sensor \mathcal{R} - proporciona información sobre la distancia radial desde el eje del haz. La coordenada z final solicitada la proporciona la posición de cada módulo en el experimento. Para ver más información sobre el sistema VELO remitirse a [21] y [12]

3.4.2. El Imán

El imán dipolo del detector LHCb permite medir la carga y el momento de las partículas. Este imán tiene la característica que puede invertir su polaridad, es decir, nos permite tomar datos cuando el campo está apuntando hacia arriba (“MagUp”) o cuando está apuntando hacia abajo (“MagDown”), lo cual nos beneficia para estudiar los efectos sistemáticos en las mediciones de asimetría.

El enorme imán del experimento LHCb consta de dos bobinas idénticas en forma de silla de montar. Ambas bobinas tienen un peso de 27 toneladas con un campo magnético de a lo mas 1.1T. En la Figura 3.10 se muestra el imán del detector LHCb. Para mayor información ver [9].

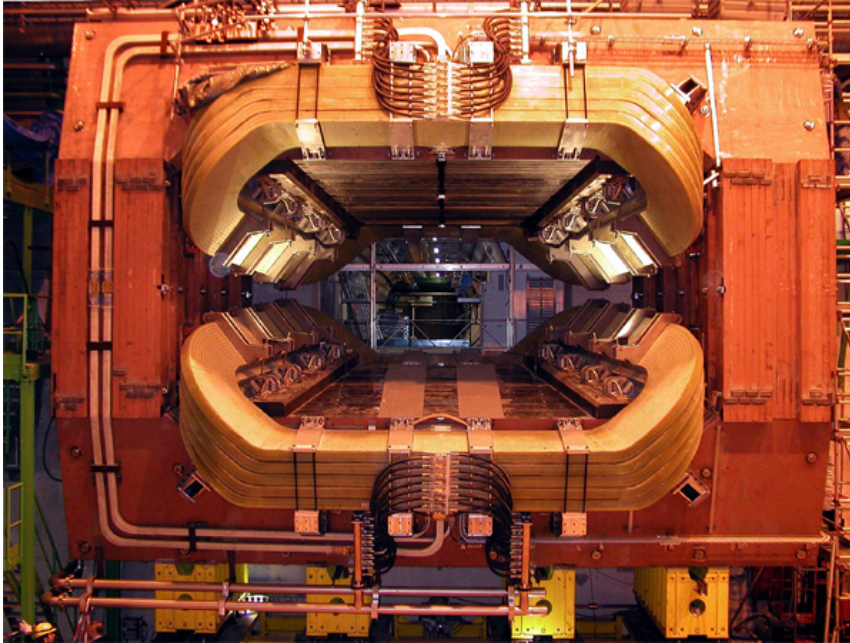


Figura 3.10: Imán dipolo del experimento LHCb

3.4.3. Sistema de rastreo

El sistema de rastreo nos ayuda a proporcionar una reconstrucción eficiente de las trayectorias de las partículas cargadas. Estos se utilizan para determinar los momentos de las partículas y a separar con precisión los vértices primarios y secundarios de las desintegraciones. Este sistema esta compuesto por el VELO (el cual vimos en la sección 3.4.1) y cuatro estaciones de seguimiento ubicadas en diferentes áreas del detector. La primera estación de seguimiento llamada Tracker Turicensis (TT) y esta ubicada entre el RICH-1 y el imán dipolo (ver Figura 3.6). Las otras tres estaciones

son T1-T3 divididas en dos regiones (región interna (IT) y región externa (OT)) están ubicadas a más de 3 metros entre el imán y RICH-2, ilustrados en la Figura 3.11.

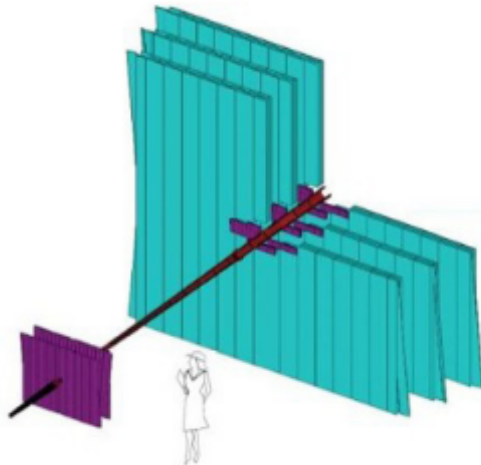


Figura 3.11: Representación de las estaciones de rastreo del detector LHCb. En la figura están presentes las estaciones de rastreo (púrpura) y el rastreador externo (verde).

Rastreador Turicensis (TT)

El rastreador de Turicensis anteriormente era conocido como Trigger Tracker o TT, utiliza detectores de microcinta de silicio con tiras largas de lectura y con pasos de tira de aproximadamente $200 \mu m$. El TT es una estación de seguimiento planar de 150 cm de ancho y 130 cm de alto que se encuentra arriba del imán dipolo del LHCb y cubre la aceptación total del experimento. El TT posee cuatro capas de sensores de silicio, que se componen de una serie de módulos. Cada capa tiene una disposición "x - u - v - x" (ver Figura 3.12), se puede observar que la primera y la última capa tienen tiras verticales, mientras que la segunda y tercera capa tiene una rotación de -5° y 5° , respectivamente. Para mayor información ver la sección 5.2 de [2].

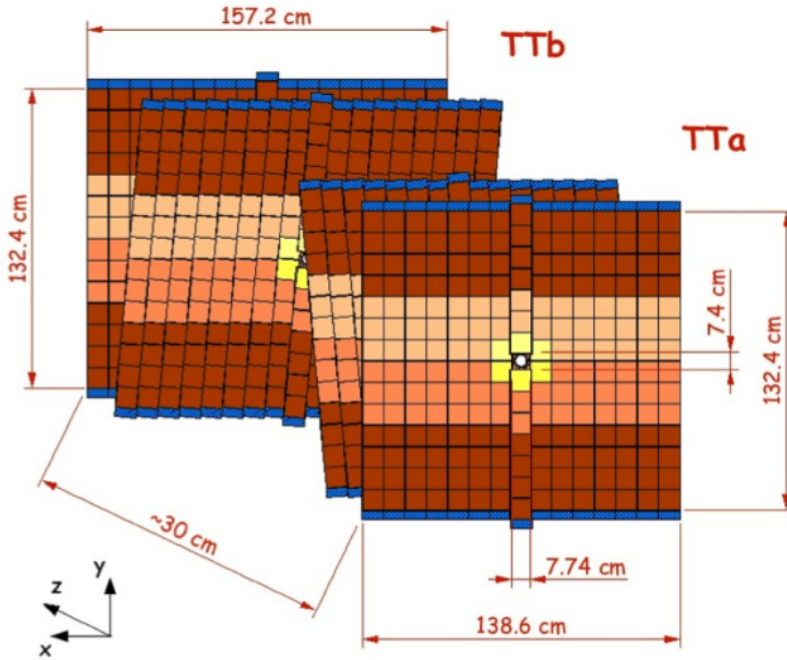


Figura 3.12: La disposición “ $x - u - v - x$ ” del TT, y las cuatro capas de las estaciones.

Rastreador interno (IT)

El sistema de Inner Tracker o IT utiliza al igual que el TT microtiras de silicio. El sistema IT está compuesto de tres estaciones T1 - T3, que están dispuestas alrededor del tubo de haz del LHC y no cubren la aceptación total del LHCb. El IT cubre una región en forma de cruz de aproximadamente 120 cm de ancho y 40 cm de alto en el centro de tres grandes estaciones abajo del imán del detector. Este sistema detecta aproximadamente el 20% de las trayectorias de partículas que pasan por las estaciones de seguimiento. .

Cada una de estas estaciones está compuesta por cuatro cajas dispues-

tas alrededor del haz y cada caja tiene cuatro capas de sensores de silicio, organizados en grupos de siete módulos. Como se muestra en la Figura 3.13. Los módulos tienen sensores ligeramente diferentes. los módulos por encima o por debajo del tubo del haz contienen un solo sensor de $320 \mu\text{m}$, mientras que los módulos laterales contienen dos sensores de $420 \mu\text{m}$. (ver [15])

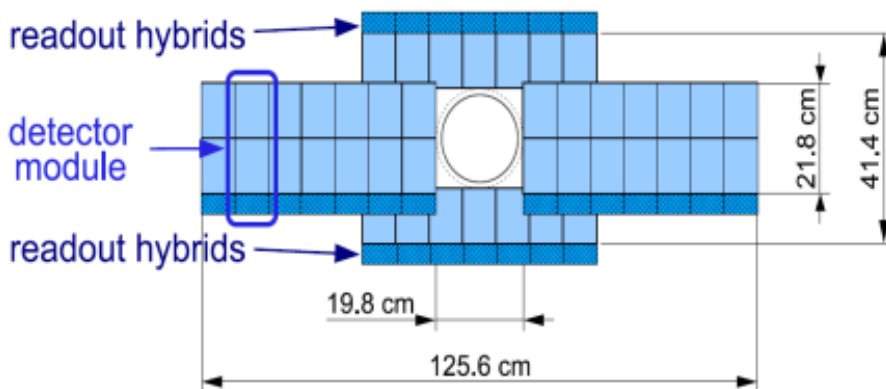


Figura 3.13: Vista frontal de los módulos detectores, que rodea el haz de luz central del LHC.

Rastreador externo (OT)

El sistema de las regiones exteriores de las estaciones de rastreo T1 - T3 es una matriz de módulos de tubos de pajilla, cada uno contiene dos capas escalonadas de tubos de deriva. Cada capa doble se compone de 4608 tubos de pajilla, dichas capas también tienen una geometría x-u-v-x como en las estaciones IT y TT. Los tubos tienen cada uno un alambre en el centro y están llenos de una mezcla de gases Ar (70 %), CO_2 (28.5 %) y O_2 (1.5 %). Cuando una partícula cargada pasa a través del tubo, el gas se ioniza y produce electrones que son acelerados hacia el alambre, y la trayectoria de la partícula está determinada por el tiempo que tardan los electrones producidos en llegar al cable (vea la Figura 3.14). El tiempo de

deriva del sistema OT es inferior a 50 ns, lo que ofrece una resolución de coordenadas de deriva de 200 μm . (ver [13])

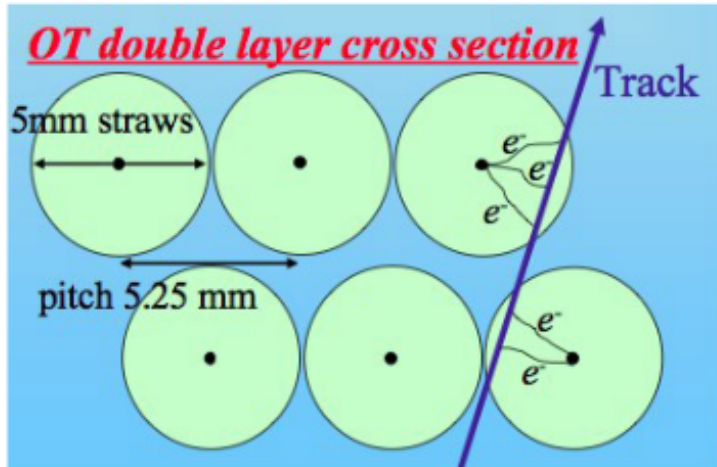


Figura 3.14: Visualización del paso de una partícula cargada a través de un módulo OT

3.4.4. Detectores Cherenkov (RICH) de imágenes anulares

Una de las características principales del experimento LHCb es la identificación de partículas como piones, kaones, protones, electrones y muones esto es crucial para reducir ruido en los estados finales de las partículas. Principalmente es de particular interés la separación correcta entre piones y kaones para el estudio de los hadrones de belleza y encanto que es uno de los objetivos primordiales del LHCb.

El sistema RICH tiene la tarea de identificar partículas cargadas en el rango de momento de 1 a 150 GeV/c , dentro de una aceptación angular de 10 a 300 milirradianes (mrad). Las partículas cargadas, cuando viajan a través de un medio dieléctrico a velocidades mayores que la velocidad de la luz en ese medio, polarizan y excitan las moléculas en el medio, que luego emiten fotones. Es así como la partícula cargada emite un cono de luz en un fenómeno conocido como radiación Cherenkov. El ángulo de emisión

del cono de radiación θ esta relacionado con la trayectoria de la partícula, dicho ángulo viene dado por:

$$\cos\theta = \frac{1}{\beta n} \quad (3.4.1)$$

donde θ es el ángulo de Cherenkov, n es el índice de refracción y β es la velocidad de la partícula. El detector LHCb contiene dos subdetectores de medición de radiación Cherenkov, RICH-1 y RICH-2 puede ver la Figura 3.15 , estos se colocan en diferentes lugares después del punto de interacción (ver Figura 3.6). La primera estación RICH-1, se encuentra entre el VELO y TT aproximadamente a 1 m del punto de interacción y la segunda estación RICH-2 se encuentra entre la estación T3 y la estación de muones, aproximadamente a 10 m del punto de interacción.



Figura 3.15: Representación esquemática de los subdetectores (izquierda) RICH-1 y (derecha) RICH-2

La primera estación RICH-1 fue proyectado para la detección de partículas con bajos momentos (2-60 GeV), que surgen de grandes ángulos polares.

Por otro lado, la estación RICH-2 se proyecta para la detección de partículas con alto momento (15-100 GeV), que surgen de ángulos pequeños. Para mayor detalle consultar [10].

3.4.5. Calorímetros

El sistema de calorímetro del LHCb está diseñado para detener las partículas a medida que pasan por el detector, midiendo la cantidad de energía perdida a medida que cada una se detiene. Los calorímetros proporcionan la principal forma de identificar partículas que no poseen carga eléctrica, como fotones y neutrones. El objetivo del sistema de calorímetro LHCb es disparar electrones, fotones y hadrones, además proporciona medidas de energía y posición de las partículas producidas en su aceptación angular.

El sistema de calorímetro del experimento LHCb está compuesto por: el calorímetro electromagnético (ECAL), el calorímetro hadrónico (HCAL) y dos detectores, el detector de almohadilla centelleante SPD y el detector previo a la ducha PD. El calorímetro electromagnético es el encargado de medir la energía de las partículas más ligeras, como electrones y fotones, mientras que el calorímetro hadrónico del experimento toma muestras de la energía de los protones, neutrones y otras partículas que contienen quarks.

Los calorímetros realizan las medidas de la energía total depositada en el medio cuando este es atravesado por una partícula. La partícula, después de una cierta distancia X_0 (longitud de onda de radiación), produce nuevas partículas con menor energía. Este efecto se llama ducha.

El SPD determina si las partículas que golpean el sistema del calorímetro están cargadas o son neutras, mientras que el PS indica el carácter electromagnético de la partícula (es decir, si es un electrón, si está cargada, o un fotón, si es neutral). Se utilizan en el nivel de activación en asociación con el ECAL para indicar la presencia de electrones, fotones y piones neutros. Ambos calorímetros tienen una estructura tipo sándwich, con capas alternas de placas de metal y plástico. Cuando las partículas golpean las placas de metal, producen lluvias de partículas secundarias. Estos, a su vez, excitan las moléculas de poliestireno dentro de las placas de plástico, que emiten luz ultravioleta. La cantidad de UV producido es proporcional

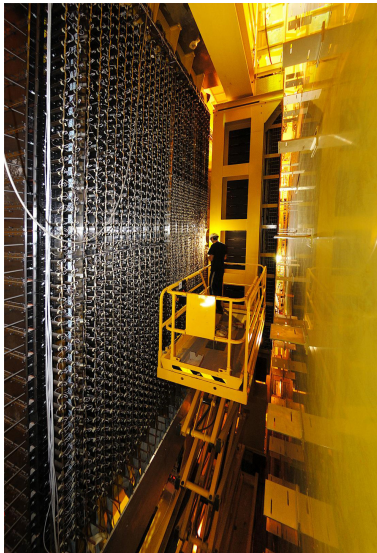


Figura 3.16: Representación esquemática

a la energía de las partículas que ingresan al calorímetro. [11].

3.4.6. Sistema de muones

Los muones son partículas pequeñas similares a los electrones, las mismas están presentes en las etapas finales de muchas desintegraciones de mesones B, por lo que la identificación de muones fuera de línea (offline) es de vital importancia para el experimento LHCb. Además, los muones son importantes para el etiquetado de sabor de quarks, y se utilizan tanto en mediciones de oscilación como de asimetría CP.

El sistema de muones está compuesto por cinco estaciones (M1-M5) de forma rectangular, que se colocan a lo largo del eje del haz (ver Figura 3.6), M1 se coloca 12,1 m después del punto de interacción y está separado de las otras cuatro estaciones por los calorímetros electromagnéticos y hadrónicos. Las otras cuatro estaciones M2-M5 se colocan al final del detector, respectivamente a 15,2 m, 16,4 m, 17,6 m y 18,8 m después del punto de interacción. A estas cuatro estaciones solamente pueden llegar

los muones, para evitar cualquier clase de ruido estas cuatro estaciones están intercaladas con filtros de hierro de 80 cm de espesor, que se utilizan precisamente para seleccionar muones. En total, el sistema está compuesto por 1380 cámaras (276 en cada estación), unos 2,5 millones de cables, suficientes para extenderse desde Ginebra hasta Madrid y cubre un área total de 435 m^2 , aproximadamente del mismo tamaño que una cancha de baloncesto.

Cada estación se divide en cuatro regiones, R1 a R4, con una distancia creciente desde el eje del haz. En la Figura 3.17 se muestra una vista esquemática lateral del sistema de muones. Todas las regiones tienen aproximadamente la misma aceptación, es decir, mantener la misma ocupación del detector en todas las regiones.

La información debe recopilarse en 20 nanosegundos, por lo que los detectores están optimizados para la velocidad. Por ende, el sistema está equipado con cámaras proporcionales de cables múltiples (MWPC), que se utiliza en todas las estaciones y el Gas Electron Multiplayer (GEM) que se utilizan en la región mas interna (R1) de la estación M1, donde la incidencia es más alta.

Cada estación contiene cámaras llenas de una combinación de tres gases: dióxido de carbono, argón y tetrafluorometano. Los muones que pasan reaccionan con esta mezcla y los electrodos de alambre detectan los resultados. Para mayor información consultar [14].

3.5. Sistema de Trigger

Recordando que las colisiones ocurren cada $25 \mu\text{s}$. El experimento LHCb funciona con una tasa de 16 MHz, lo cual es menos de la mitad de la tasa del cruce del haz de protones del LHC que es de 40 MHz. Aún así esta tasa de colisión mas pequeña es mas grande de lo que el hardware del detector es capaz de procesar y almacenar para análisis posteriores, esto implica que se pierdan eventos significativos. Es por este motivo que el detector LHCb tiene que tener un sistema eficiente que nos permita seleccionar únicamente datos de interés, es decir, un primer gran filtro dado que la capacidad de almacenamiento es finito y desde ya la cantidad de datos que se guardan al año es de 50 peta-bytes, es una cantidad increíble de datos. En conclusión, el almacenamiento es muy costoso y por eso necesitamos de un primer gran

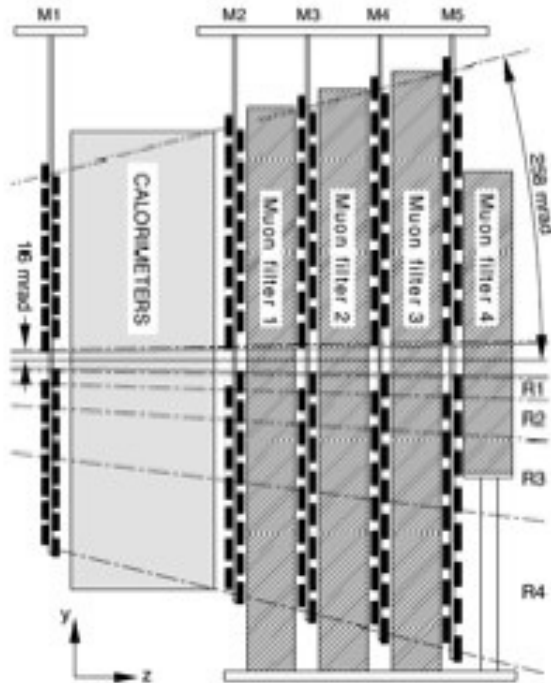


Figura 3.17: Vista lateral del sistema de muones de M1-M5. También se muestran las cuatro regiones R1-R4

filtro.

Esta selección de eventos se optimiza utilizando tanto la topología de las desintegraciones del mesón B , cuya vida útil es larga, lo cual, nos permite a formar vértices de desintegración a partir del punto de interacción, y las partículas con una p_T alta, debido a la gran masa invariante del mesón B .

Ese sistema de selección de eventos se llama *Trigger*, el cual tiene dos niveles el trigger de nivel-0 (L0) y el trigger de alto nivel (HLT), este a su vez se divide en dos subniveles (HLT1 y HLT2). La representación de estos niveles se puede observar en la Figura 3.18

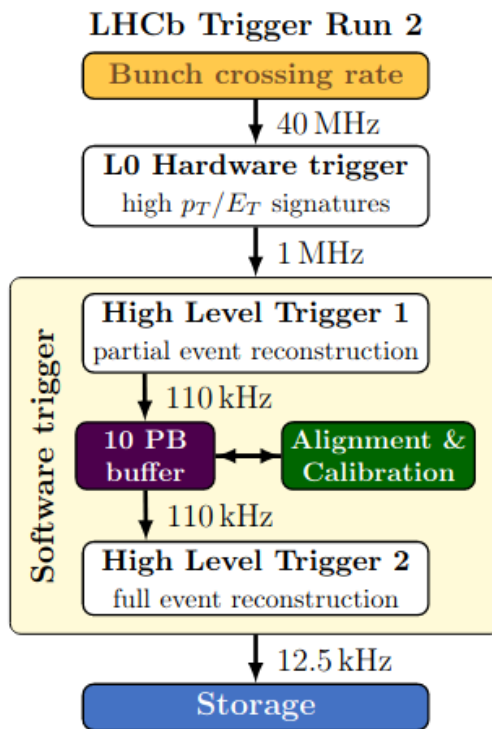


Figura 3.18: Representación de los niveles del Sistema Trigger (Imagen tomada de [17])

3.5.1. Nivel-0 (L0)

El nivel-0 **L0** o también llamado nivel de **Hardware** utiliza la información obtenida por los calorímetros electromagnéticos y hadrónicos, la cámara de muón, el magneto y la cámara de trazos. y lo que busca rápidamente es seleccionar eventos que puedan representar electrones, fotones y hadrones con alta energía transversal y muones de alto momento transversal. La información que proporcionan se combina en la unidad de decisión de nivel 0 (L0DU), que luego es responsable de la decisión final sobre si aceptar o no un evento, esta decisión se toma $4 \mu m$ después del cruce del haz. La velocidad de cruce de aproximadamente 16 MHz utilizada por el

experimento LHCb se reduce a aproximadamente 1 MHz con este primer nivel de activación.

Aún al haber pasado por la selección de L0 queda mucho ruido, por lo tanto, es necesario para los analistas hacer una selección **offline**, es decir, una **selección mas refinada de los datos**. En esta parte offline es donde este trabajo va a tener su mayor repercusión o validez porque se va a refinar la selección de datos para un tipo de física.

3.5.2. Trigger de alto nivel (HLT)

El segundo nivel de Trigger es Trigger alto nivel, o HLT, y su función primordial es reducir aún más la tasa de transferencia de datos que pasa el trigger L0, para alcanzar una tasa de eventos lo más pequeña posible como para almacenar los eventos. La salida de nivel L0 de 1 MHz se reduce a aproximadamente 3 KHz por el HLT. El HLT opera a nivel de software, los datos son enviados a una Granja de Filtros de Eventos (EFF), donde miles de Computadoras Personales (PCs) ejecutan los algoritmos del HLT, programados en C++. Al estar basado en software, el HLT es muy flexible y se puede actualizar con el tiempo.

El HLT se divide en dos etapas denominadas HLT1 y HLT2, que tienen diferentes propósitos. La primera etapa HLT1 se hace una reconstrucción parcial de eventos, esta busca partículas con alto P_T las cuales no se originan en el vértice primario. HLT2 se realiza la reconstrucción total de los eventos y se hace una selección de datos exclusiva e inclusiva. En esta etapa se utilizan algoritmos más rigurosos como redes neuronales y algoritmos basados en la topología de los decaimientos de dos, tres o cuatro cuerpos.

La información se manda a distintos países, cuando nosotros accedemos a recoger estos datos los podemos estar jalando de cualquier parte del mundo entonces es una red computacional científica global. Esta red es necesaria porque la cantidad de datos almacenada es gigantesca. Para mayor información del sistema de Trigger consultar [17] y [16].

3.6. Sistemas de software

Todos los sistemas de software del LHCb [24] se basan en el marco de Gaudi, todos los sistemas comparten y se comunican a través del modelo de evento LHCb, como se muestra en la Figura 3.19 y tienen la capacidad de reconstruir (Brunel) y simular (Gauss) eventos, así como el análisis de datos (DaVinci). Los sistemas se describen a continuación:

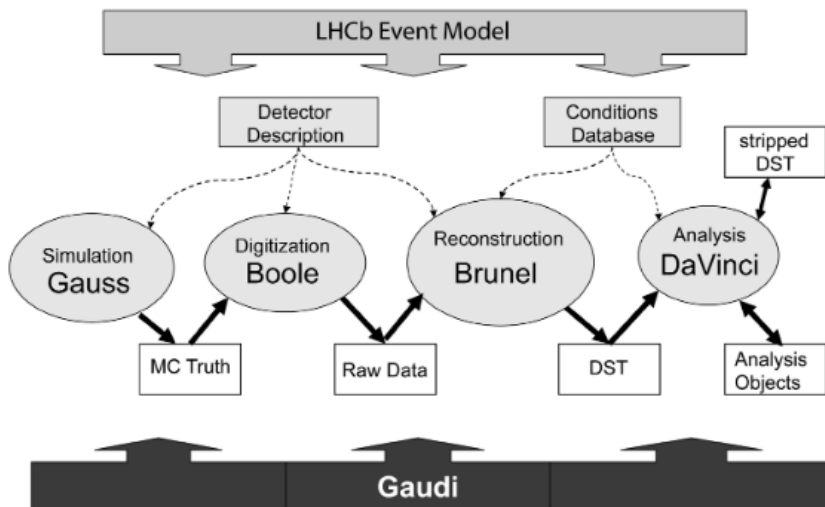


Figura 3.19: Representación de los sistemas de software de procesamiento de datos del LHCb, [24]

1. **Gauss:** El sistema Gauss simula lo que sucederá en el espectrómetro del LHCb y con ello ayuda a comprender las condiciones experimentales y el rendimiento, así como proporciona las simulaciones de Monte Carlo de canales de decaimiento de interés. Lo cual es muy importante en los análisis de decaimiento, en esta tesis ha sido esencial. Este sistema integra dos fases independientes que utiliza tres aplicaciones: Pythia(ver), EvtGen(ver) y Geant(ver).

La primera fase consiste en la generación de eventos de las colisiones protón-protón (Pythia) y la descomposición de los mesones B (EvtGen). La segunda fase consiste en el seguimiento de las partículas,

encargada la aplicación Geant. Al final de las dos etapas nos brinda información simulada, que intenta “imitar” la producción de partículas y el rendimiento del detector LHCb.

2. **Brunel:** Es el sistema que se dedica a la reconstrucción de eventos del LHCb, procesa datos de la salida de los sistemas de simulación e inclusive datos reales. Produce archivos de salida que contienen información sobre el momento, trayectoria de las partículas, así como la identificación de partículas.
3. **DaVinci:** Es el software capaz de hacer selecciones de eventos, reconstrucciones de vértices. La información se guarda en archivos de salida de datos como ser ntuples, árboles e histogramas.

Capítulo 4

Análisis de los decaimientos

$$B^+ \rightarrow \pi^+ \pi^+ \pi^- \text{ y}$$

$$B^+ \rightarrow K^+ K^- \pi^+$$

En este capítulo se dará una descripción del conjunto de datos y simulación en la sección 4.1. En la sección 4.2 se definen las variables de interés. En las secciones 4.3 y 4.3.1 se detallada el proceso de selección de datos. Luego en la sección 4.4 se presenta la distribución del mesón B y los tipos de ruido. Seguidamente en la sección 4.5 se describe la selección offline de datos que es clave para este análisis. Finalizando en la sección 4.6 donde es descrita la optimización.

4.1. Conjunto de Datos y Simulación

En la Figura 3.5 nos muestra la luminosidad integrada colectada desde 2010 - 2018, en este estudio se utilizó parte de la muestra de datos recopilada por el LHCb en el Run 2, los cuales comprenden los años 2015 y 2016 a una energía del centro de masa de $\sqrt{s} = 13$ TeV y una luminosidad integrada de 1.9 fb^{-1} , podemos observar que la cantidad de datos del 2016 aproximadamente es el doble en comparación al 2015.

También se utilizaron muestras de simulación de Monte Carlo (MC) del

CAPÍTULO 4. ANÁLISIS DE LOS DECAIMIENTOS $B^+ \rightarrow \pi^+\pi^+\pi^-$ Y $B^+ \rightarrow K^+K^-\pi^+$

año 2015 y 2016 las cuales reproducen al máximo las condiciones reales de las colisiones, son muestras totalmente limpias sin ningún tipo de ruido, es decir, representan solamente la señal. Antes de utilizar estas muestras, las mismas pasan por el proceso de selección, descrito en la sección 4.3.1. En la Tabla 4.1 nos da un resumen de estas estadísticas.

Año	Polaridad del imán	$B^+ \rightarrow \pi^+\pi^+\pi^-$	$B^+ \rightarrow K^+K^-\pi^+$
2015	MagDown	953593	1705394
	MagUp	912365	1681936
2016	MagDown	2194223	4052663
	MagUp	2039809	4037161

Tabla 4.1: Estadísticas de señales de Monte Carlo (MC)

A las muestras de Monte Carlo se les aplica un corte basado en tablas de verdad (MCTruth o Trueid). Esto se debe a que el software de DaVinci, plataforma privada del experimento, no sabe identificar específicamente quien es la primera y segunda partícula, por lo tanto, acepta tanto piones como kaones, consecuencia de ello es que se hace un doble conteo en la muestra de señal y es por esta razón que se aplican dichos cortes, es un corte de preselección específico para cada canal de decaimiento.

Trabajar con la muestra de datos recopilada por el LHCb en el Run 2, los cuales comprenden los años 2015 y 2016, resultó un gran desafío ya que la conexión se realizaba a través de un servidor remoto del Centro Brasileiro de Pesquisas Físicas (CBPF) para acceder a los datos y realizar los análisis. Trabajar de esta forma en cierto sentido, tomó mas tiempo de ejecución en los algoritmos implementados en esta disertación.

4.2. Variables de interés

Las variables de interés se basan en características topológicas de los mesones B . En general, para la selección se esta interesado en variables que sean buenas discriminantes entre señal y ruido. Esto se logra explorando las características topológicas de las variables. En la Figura 4.1 se muestra la topología de las variables del decaimiento $B^+ \rightarrow h^+h'+h''^-$.

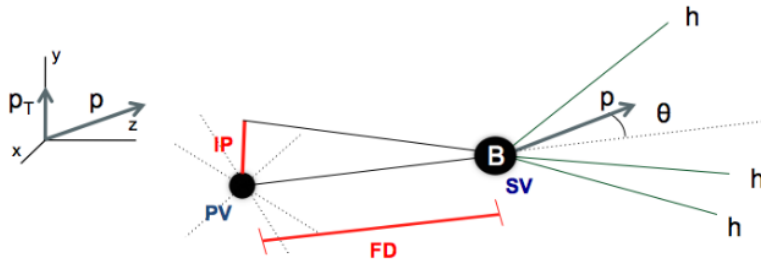


Figura 4.1: Topología de las variables

El mesón B que es un mesón pesado, se origina en el punto de interacción del vértice primario (PV), luego el mesón viaja a una distancia antes de decaer, esa distancia se denomina distancia de vuelo (FD), luego el mesón decae, justo donde decae se llama vértice secundario (SV) y decaen las tres partículas hijas. Entre las variables más utilizadas para discriminar entre eventos de señal y eventos de ruido tenemos las siguientes:

- **Masa invariante:** La masa invariante del candidato B se calcula utilizando medidas de energía y momentos de partículas de estado final identificadas. La masa invariante se define así:

$$M = \sqrt{E^2 - (\vec{p})^2} \quad (4.2.1)$$

donde, $\vec{p} = \vec{p}_1 + \vec{p}_2 + \vec{p}_3$ es la suma de los 3-momentos hijas y $E = E_1 + E_2 + E_3$ es la suma de sus respectivas energías.

- **Vértice primario y secundario (PV-SV):** El vértice primario (PV) de una partícula es donde se origina la partícula, es decir, es el punto de interacción de la colisión protón-protón. El vértice secundario (SV) es el punto donde la partícula decae.
- **Momento transverso, (\mathbf{P}_T):** Es el momento perpendicular a la línea que une el vértice primario y el vértice secundario, es decir, perpendicular a la línea de vuelo del mesón B

- **FD** : Distancia de vuelo del vértice primario donde se origina las partículas al vértice secundario donde decaen las partículas en su estado final.
- **Parámetro de impacto (IP)**: se define como la máxima aproximación de la trayectoria y el vértice primario. La variable IP nos ayuda a distinguir partículas que se originaron en el vértice primario y partículas producidas en el vértice secundario.
- **FD χ^2** : es el ajuste obtenido de la distancia de vuelo.
- **IP χ^2** : El chi-cuadrado del impacto de las madres y de las hijas
- **χ^2 del vértice secundario**: Para saber si es exactamente el vértice de donde nacieron las tres hijas y se va ajustando. El mejor χ^2 es una variable dedicada a la identificación de π y k
- **$\cos \theta$** : Coseno del ángulo θ entre el vector de vuelo candidato B y su momento.
- **Distancia de aproximación más cercana (DOCA)**: es la distancia más corta entre dos pares de trayectorias desde el estado final de las partículas.
- **Variables de identificación de partículas (PID)**: Los subdetectores de calorímetros, RICH y sistema de muones del detector LHCb, se combinan y nos brindan información sobre un conjunto único de variables PID. Estas variables PID son utilizadas como criterios de selección en los análisis para identificar piones, kaones, electrones y muones. Ahora bien, en esta disertación se emplean las variables ProbNN para reducir el ruido de mis-identificación de $K - \pi$ y $\pi - K$.

4.3. Selección

El objetivo de este estudio que representa un proceso físico de interés como fue mencionado (ver sección 2.1) es el aislamiento de una señal. Esto se logra a través del proceso de selección, que es uno de los pasos más importantes en cualquier análisis de física de alta energía basado en muestras

de datos. Estas muestras contienen diferentes tipos de ruido, por ejemplo, las muestras de datos que se utilizan en el análisis no son totalmente específicas de un canal de desintegración, es decir, que en la misma muestra de datos se pueden incluir muchos canales con el mismo estado final.

El objetivo de este proceso de selección es reducir al máximo los eventos que no provengan del decaimiento deseado, esto es a lo que llamamos ruido, pero siempre con el compromiso de mantener la mayor cantidad de eventos de señal deseados. Sin embargo, la mayoría de los métodos que se enfocan en eliminar el ruido, siempre afectan los eventos de señal. Es por este motivo, que el proceso de selección busca maximizar tres cantidades: la pureza, la eficiencia y la significancia estadística, discutiremos sobre ellas en la sección de optimización 4.6.

4.3.1. Proceso de Selección

La selección de datos para un decaimiento de tres cuerpos $B^+ \rightarrow h^+h'^+h''^-$ es un proceso que conlleva varios pasos. El seguimiento de estos pasos esta representado en la Figura 4.2.



Figura 4.2: Proceso de Selección

El sistema de Trigger ([17] y [16]) es el primer gran filtro por el cual

CAPÍTULO 4. ANÁLISIS DE LOS DECAIMIENTOS $B^+ \rightarrow \pi^+\pi^+\pi^-$ Y $B^+ \rightarrow K^+K^-\pi^+$

pasan los datos de la colisión protón - protón, es descrita en la sección 3.5. Luego pasan una selección llamada de Stripping [38], aquí los datos almacenados se procesan aún mas para separar los eventos por la física de interés, es decir, son un conjunto de cortes según la física que se desea estudiar, son cortes bien relajados basados en la cinemática del decaimiento para no afectar de ninguna forma nuestra señal pero si que permita diferenciar otros canales que no se tiene interés en analizar. Algunos cortes aplicados en la selección de Stripping están dados en la Tabla 4.2

Variable	Criterio de selección
Distancia de vuelo del χ^2	B_FDCHI2_OWNPV > 500
χ^2 del vértice secundario	B_ENDVERTEX_CHI2 < 12.
IP χ^2 del candidato B	B_IPCHI2_OWNPV < 10.
$\cos \theta$ del candidato B	B_DIRA_OWNPV > 0,99998
Máximo DOCA	B_AMAXDOCA < 0,2
P_T del candidato B	B_PT > 1000.
Suma P_T de las trayectorias	ptsum > 4500.
Suma de P de las trayectorias	psum > 20000
Suma de IP χ^2 de las trayectorias	ipchi2toppvsum > 500.

Tabla 4.2: Algunos criterios de selección de stripping 20

Ahora bien, la muestra de datos colectada por el detector contiene la familia de decaimientos $B^+ \rightarrow h^+h'^+h''^-$, siendo así, en la preselección se realiza un conjunto de cortes específicos para cada canal de decaimiento, con el objetivo de reducir el ruido físico. En este punto las muestras de datos no han tenido ningún requisito de identificación de partículas del estado final, lo cual conduce a una identificación errónea de piones y kaones llamado ruido de alimentación cruzada de otros canales $B^+ \rightarrow h^+h'^+h''^-$. Para reducir este tipo de ruido se aplican cortes flexibles a las variables PID de los dos canales de decaimiento, estas variables son ProbNNK que da el valor de probabilidad de que la partícula sea un Kaón (K) y ProbNNpi que da el valor de probabilidad de que la partícula sea un pión (π). En las Figuras 4.3 y 4.4 se muestra la distribución de probabilidad de estas variables para d1, d2 y d3 que son las partículas hijas. Por ejemplo, d1_ProbNNK es la probabilidad de que la primera partícula hija sea un kaón.

CAPÍTULO 4. ANÁLISIS DE LOS DECAIMIENTOS $B^+ \rightarrow \pi^+\pi^+\pi^-$ Y $B^+ \rightarrow K^+K^-\pi^+$

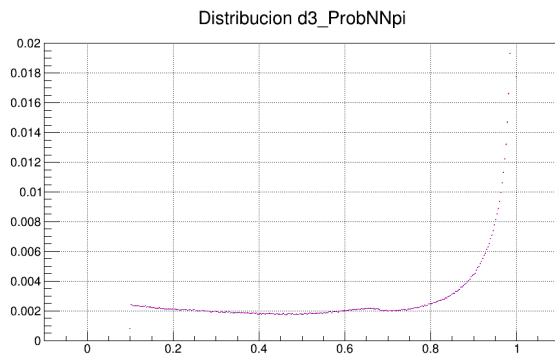
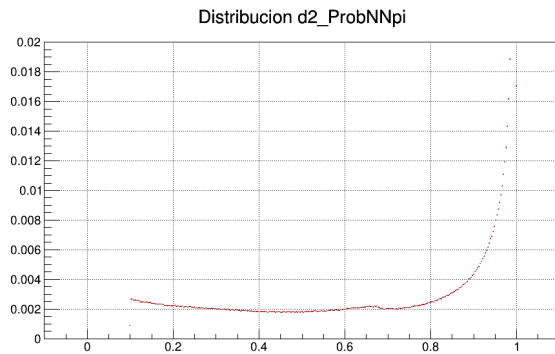
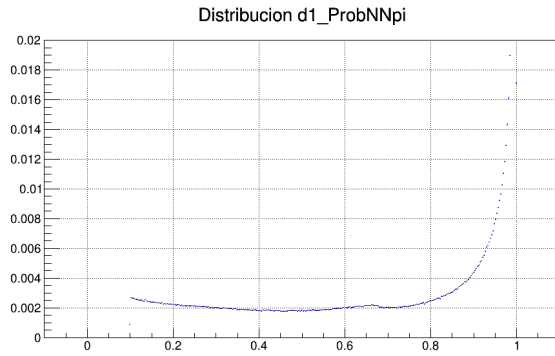
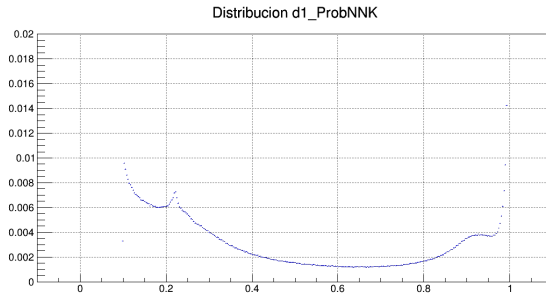
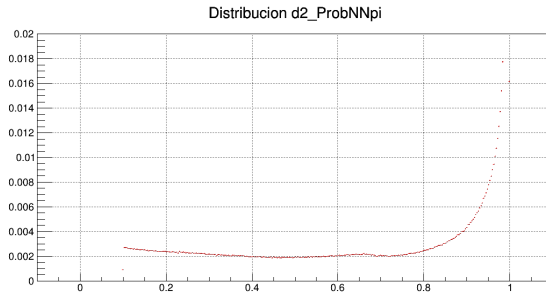


Figura 4.3: Distribución de probabilidad de, (a): d1_ProbNNpi, (b): d2_ProbNNpi y (c): d3_ProbNNpi de $B^+ \rightarrow \pi^+\pi^+\pi^-$

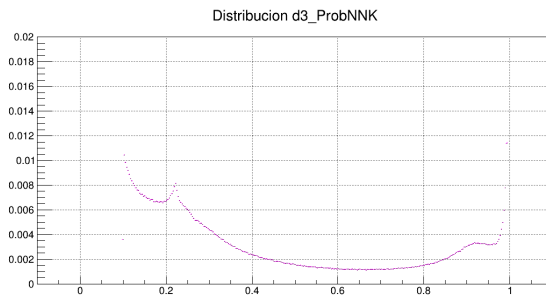
CAPÍTULO 4. ANÁLISIS DE LOS DECAIMIENTOS $B^+ \rightarrow \pi^+\pi^+\pi^-$ Y $B^+ \rightarrow K^+K^-\pi^+$



(a)



(b)



(c)

Figura 4.4: Distribución de probabilidad de, (a): d1_ProbNNK, (b): d2_ProbNNpi y (c): d3_ProbNNK de $B^+ \rightarrow K^+K^-\pi^+$

Los cortes mínimos de PID aplicados a las partículas de kaones y piones son, $\text{ProbNNK} > 0,1$ y $\text{ProbNNpi} > 0,1$, se deshace de los picos que se observan (Figura 4.4) que son ruidos para las muestras. Esto permite tener una distribución de masa invariante mas limpia como punto de inicio de análisis. Claro, más estudios necesitan ser hechos que es lo que se explica más adelante con la selección de identificación de partículas.

4.4. Distribución del Mesón B

Se estudiarán los canales $B^+ \rightarrow \pi^+\pi^+\pi^-$ y $B^+ \rightarrow K^+K^-\pi^+$, después de haber aplicado los tres primeros pasos del proceso de selección de eventos en la muestra de datos, se procederá a seleccionar un corte con la ayuda del análisis multivariante, dicho corte ayudará a disminuir la cantidad de ruido blanco o ruido combinatorial posible discutidos en la sección 4.4.1. En la Figura 4.5 se aprecia la distribución de la masa invariante (B_m) para el canal $B^+ \rightarrow K^+K^-\pi^+$ con los datos del 2015 y 2016, como se observa a pesar de la preselección, existe un alto nivel de ruido, aun así se puede visualizar una leve pretuberancia en torno a la masa del mesón B.

La Figura 4.5 representa la distribución de masa invariante del mesón B y no esta normalizado porque se contabiliza todos los eventos que fueron registrados por el detector LHCb, es decir, es toda la estadística colectada en los años 2015 y 2016 que son los estudiados en esta disertación. En cada evento se pretende que haya solo un candidato al mesón B decayendo en $B^+ \rightarrow K^+K^-\pi^+$, entonces por naturaleza este no es un gráfico de probabilidad.

Cuando la distribución de la variable es Gaussiana nos da la confianza de que esa distribución pertenece a ese canal de decaimiento. Si vemos nuestro histograma inicial (Figura 4.5) no es una Gaussiana limpia como esperábamos. A pesar que los datos han pasado por varios filtros basados en la física y en la topología del decaimiento. Es un proceso que tiene muchas otras partículas o contaminación de otras partículas, es decir, tiene ruido.

Es una distribución que no es simétrica a ambos lados como se esperaría la señal para este decaimiento, la masa de B es 5279 MeV y vemos que hay un pico aproximadamente en 5279 pero es una muestra que esta altamente

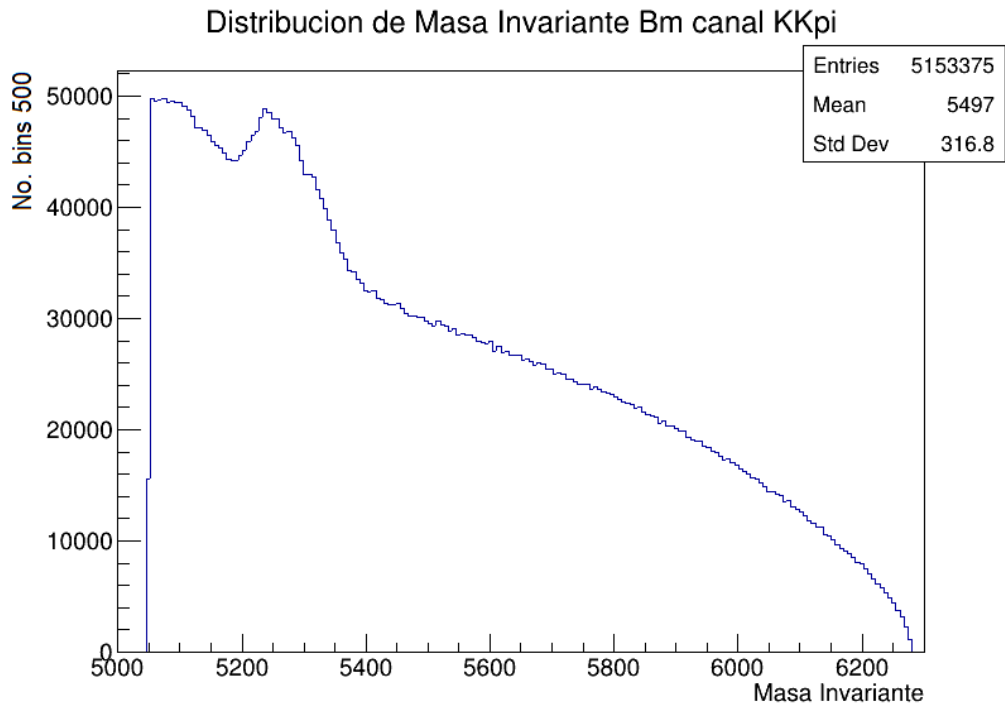


Figura 4.5: Distribución del Mesón B, canal $B^+ \rightarrow K^+K^-\pi^+$

contaminada y con grandes niveles de ruido, se puede observar que la distribución del tipo de ruido a la derecha es diferente del tipo de ruido a la izquierda. *Como primer paso queremos disminuir al máximo el ruido que esta a la derecha, el cual podemos decir que es ruido combinatorial.* Se discutirá sobre los distintos tipos de ruido en la siguiente sección.

4.4.1. Tipos de ruido

Como se discutió en la sección 4.4, cuando la distribución no sigue una distribución Gaussiana se sospecha que esa distribución no pertenece a ese canal, lo cual es producto de diferentes tipos de ruido. Estos tipos de ruido logran pasar los criterios de selección de Trigger y Stripping e ingresan a la región de señal debido a varias razones. Para el decaimiento en estudio

$B^+ \rightarrow h^+h'^+h''^-$ las principales fuentes dominantes de ruido son:

1. **Ruido Combinatorial:** Es el resultado de la asociación aleatoria de tres trazos no relacionados que aparentan ser el decaimiento en estudio $B^+ \rightarrow h^+h'^+h''^-$. Esto significa que cuando se hizo la construcción un trazo que no tenía nada que ver con el decaimiento, otro trazo igual que no tenía nada que ver con el decaimiento y otro trazo en las mismas circunstancias coincidieron en un punto y se seleccionaron como si fuera el decaimiento en estudio pero realmente ellos no nacieron de ese punto, entonces son simplemente trazos aleatorios. La ventaja de este ruido es que como son trazos totalmente descorrelacionados en general son ruidos bien portados.
2. **mis-identificación:** Este ruido se debe a los decaimientos de tres cuerpos $B^+ \rightarrow h^+h'^+h''^-$ que se reconstruyen con partículas mal identificadas, es decir, este error se produce al identificar una partícula π como un K en su estado final. Este error de identificación de partículas conduce a un pico de masa más bajo o más alto que la masa de B. Por ejemplo cuando un π se identifica como un K esto conduce a un pico de masa más bajo dado que la masa del π es menor que la masa del K y esto nos lleva a que la distribución del mesón B no tenga el comportamiento de una Gaussiana.
3. **Decaimientos parcialmente reconstruidos de cuatro cuerpos:** Este tipo de ruido ocurre cuando tenemos un decaimiento en cuatro cuerpos y una de las partículas no se reconstruyó entonces este decaimiento de 4 cuerpos se confundió con uno de tres cuerpos y como no representa en sí mi decaimiento su masa va a quedar fuera de lo que se denomina mi región de señal. Este tipo de ruido se modela en el ajuste de masa invariante B .

Todos estos tipos de ruido pueden quedar abajo de la región de señal, dentro o alrededor de ella (ver Figura 4.6), como primer paso en este estudio es disminuir al máximo el ruido combinatorial para ello solo se utilizará la región de la masa invariante arriba de 5400 MeV ($B_m > 5400$), área roja señalada en la Figura 4.6, la región verde contiene señal mas ruido, aquí predomina el ruido de mis-identificación, es por tal motivo que esa área es

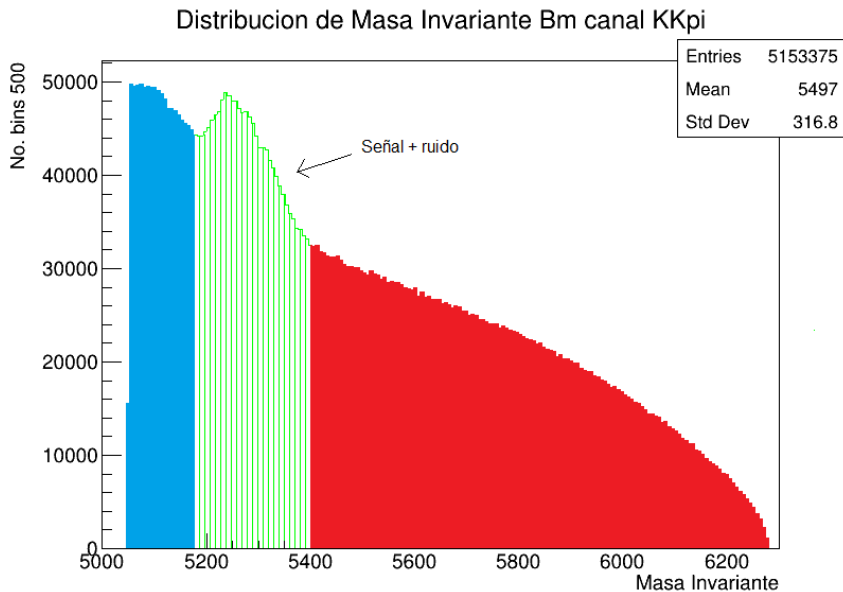


Figura 4.6: Distribución del Mesón B: área roja ($B_m > 5400$) ruido combinatorial, área verde: señal mas ruido de distintos tipos, área azul: predominante los decaimientos parcialmente reconstruidos de cuatro cuerpos.

altamente desafiante. Ese será el último paso eliminar la mayor cantidad de eventos de ruido con el compromiso de mantener nuestra señal.

Simplemente cortar el área de ruido, por ejemplo, el área roja no es una solución óptima al problema, se puede decir que sería una opción, sin embargo, para estudios posteriores como ser ajuste de masa modelar la función de distribución de probabilidad (pdf) resultante es muy difícil. De igual forma, aunque se corte siempre se tendrán eventos de ruido abajo de la señal. Por lo tanto entender estos tipos de ruido permiten extrapolar cual es la contaminación de ellos en la región de señal. Al final todo esto es para disminuir al máximo los eventos de ruido abajo de la señal, para obtener una pdf que sea fácil de modelar con el ajuste de masa.

4.5. Selección Offline

Esta selección offline involucra los últimos dos pasos del proceso de selección descrito en la sección 4.3.1, la selección de análisis multivariante (MVA) y la selección de identificación de partículas (PID), son los que tienen mayor relevancia en este estudio, estos buscan eliminar las dos principales contribuciones de ruido que afectan la señal. El ruido combinatorial y el ruido debido otros canales hadrónicos.

4.5.1. Selección Análisis Multivariante

El método que se utilizará para clasificar las partículas con el objetivo de reducir el ruido combinatorial es mediante técnicas análisis multivariante (MVA), la cual se describe en la sección 5.3, en este estudio se utilizará el algoritmo de Boosted Decision Tree (BDT) descrito en la sección 5.4.1 para encontrar ese corte óptimo de BDT, esto se lograra mediante la optimización de las denominadas figuras de merito como serán definidas posteriormente.

4.5.2. Selección de Identificación de Partículas (PID)

Después de los criterios de selección por los que han pasado los datos, los cuales son Trigger, Stripping, preselección y selección de análisis multivariante (enfocado a reducir el ruido combinatorial). El ruido restante es debido a una identificación errónea de piones o kaones llamada Misidentificación estos forman el de contaminación hadrónica en 3-cuerpos y el ruido de decaimientos parcialmente reconstruidos discutidos en la sección 4.4.1. Para eliminar estos tipos de ruido restante se utiliza el análisis de identificación de partículas (PID), el cual juega un papel clave en el entorno de selección. Esta selección PID al igual que la selección de MVA, busca eliminar el ruido manteniendo la mayor parte de la eficiencia de la señal. La selección de PID que se encontrará debe ser mas fuerte que los cortes flexibles de PID ya aplicados en la preselección.

4.6. Optimización

La optimización se logra encontrando el corte óptimo en la salida de BDT, este corte tiene la finalidad de reducir el ruido combinatorial pero con el compromiso de mantener la mayor eficiencia de la señal. El corte se encuentra agregando la variable BDT a las muestras y recorriendo todos los posibles cortes de BDT, así encontramos el corte óptimo que maximice la **figura de mérito (FoM)**. La figura de mérito es una cantidad que es función de una serie de parámetros, por ejemplo la significancia estadística y el número de eventos, el gran objetivo será maximizar esta cantidad. A continuación definiremos algunas cantidades que se pueden maximizar:

- **Significancia Estadística:**

$$Significancia = \frac{S_{MC}}{\sqrt{(S+B)_{datos}}} \quad (4.6.1)$$

donde, S_{MC} es el número de eventos tomados de las muestras de señal de Monte Carlo y $(S+B)_{datos}$ es el número de eventos en la región de señal definida por $|B_m - 5279| < 40 MeV/c^2$, tomada de los datos reales con los cortes de preselección. Esta definición es diferente a la significancia estadística en el contexto de pruebas de hipótesis.

- **Pureza:**

$$Pureza = \frac{S_{MC}}{(S+B)_{datos}} \quad (4.6.2)$$

donde, S_{MC} y $(S+B)_{datos}$ fueron definidos en la significancia estadística 4.6.1

- **Eficiencia del corte BDT:**

$$Eficiencia = \frac{S_{corte}}{S_{Total}} \quad (4.6.3)$$

donde, S_{corte} es el número de eventos de señal que pasan el corte y S_{Total} es el número total de eventos de señal antes que se realice algún

corte. Estas eficiencias del corte se determinan mediante muestras de simulación de señal (MC). Podemos decir, que la eficiencia de BDT va a ir disminuyendo a medida aumente el valor del corte.

■ Eficiencias PID

Las variables de identificación de partículas (PID) no están bien descritas en las simulaciones de Monte Carlo (MC), variando de las que se encuentran en los datos, por lo tanto, estas eficiencias se logran obtener por medio de un paquete llamado **PIDCalib** [30] que significa “Calibración de Identificación de Partículas” implementado en ROOT. Este paquete es un conjunto de herramientas para ayudar a los analistas a calcular la eficiencia de los criterios de selección de identificación de partículas (PID).

Mediante estas eficiencias estimadas para muestras simuladas se encontrará el criterio de selección PID para las variables ProbNN, el cual es el ultimo paso en el proceso de selección. Dichas eficiencias se logran obtener como peso para cada polaridad de imán y cada trayectoria del estado final, estos pesos se van agregando evento por evento en la región de aceptación. Los pasos para la implementación del paquete PIDCalib están descritos en [1].

El proceso de selección busca optimizar la pureza, la eficiencia del corte y la significancia estadística, no obstante, suele ser bien complicado maximizar varias cantidades simultáneamente, por esta razón vamos a maximizar la **significancia estadística** que nos brinda un equilibrio óptimo entre la pureza y la eficiencia del corte.

La optimización de estas cantidades es una etapa crucial de este análisis.

Capítulo 5

Machine Learning

El machine learning (ML) es una rama de la inteligencia artificial que tiene como objetivo desarrollar algoritmos que permitan a la máquina aprender de la experiencia. Este capítulo nos dará información sobre que es machine learning en la ciencia de datos en la sección 5.1. La sección 5.2 hace referencia a los enfoques de la clasificación en ML. La sección 5.3 detalla la motivación para utilizar métodos multivariados. Los algoritmos utilizados de ML son descritos en la sección 5.4. Finalizando en la sección 5.5 con la evaluación del desempeño de la clasificación.

5.1. Machine Learning en la Ciencia de Datos

El término de machine learning (ML) nació a principios de los años 50, no es fácil encontrar una definición, sin embargo, diremos que machine learning es una sub-disciplina de la inteligencia artificial cuyos algoritmos tienen la capacidad de aprender y tienen la capacidad de resolver problemas sin ser programados explícitamente para resolver esos problemas, ellos aprenden de los datos, logran inferir de los datos reglas para solucionar el problema.

También podemos ver a machine learning como la inferencia estadística llevada al computador, hay una mezcla entre dos campos de la ciencia de datos (ciencias de la computación y estadística) que es lo que da lugar a

machine learning. Y se trata de encontrar el modelo que mejor describa a los datos.

En ML hay dos paradigmas más utilizados dependiendo de la naturaleza del aprendizaje:

1. Aprendizaje supervisado
2. Aprendizaje no supervisado

En el aprendizaje supervisado al programa se le proporcionan los datos de entrada (o inputs) de los que tiene que aprender asociados con las respuestas (o outputs) que para ellos se esperan del programa. A este conjunto de datos se le llama conjunto de entrenamiento (training). El programa se encarga de alterar sus parámetros para que la respuesta sea óptima. Dependiendo de si la respuesta es bien un número, bien una etiqueta (label) o clase, será un problema de **regresión** o de **clasificación**. En el caso del aprendizaje no supervisado al programa no se le proporcionan las respuestas deseadas, sino solamente los datos de entrada, y él debe encontrar algún tipo de estructura en ellos, detectando por ejemplo, características desconocidas. Un ejemplo de ellos son los problemas de agrupación o clustering [27]

El objetivo del aprendizaje supervisado es hacer inferencia estadística de la distribución que generan los datos y predecir el output “ y_{new} ” dada una nuevo input “ x_{new} ”.

En ML tenemos dos categorías de datos diferentes, los *datos de entrenamiento* con los que el algoritmo va a aprender y hacer el modelo y los *datos de evaluación o test* con los cuales se va a evaluar el modelo para ver si aprendió y no hizo un sobre ajuste “overfitting”.

El funcionamiento del entrenamiento en los algoritmos de machine learning se puede visualizar en la Figura 5.1. El objetivo es hacer una predicción y lo compara con los resultados esperados, dependiendo que tan bien le haya ido, el algoritmo va a hacer una retroalimentación actualizando los parámetros, esto es un ciclo. Luego que el algoritmo esta entrenado tenemos un modelo de ML, al cual le pasamos los datos de evaluación para ver si logra predecir bien sin caer en un sobre ajuste.

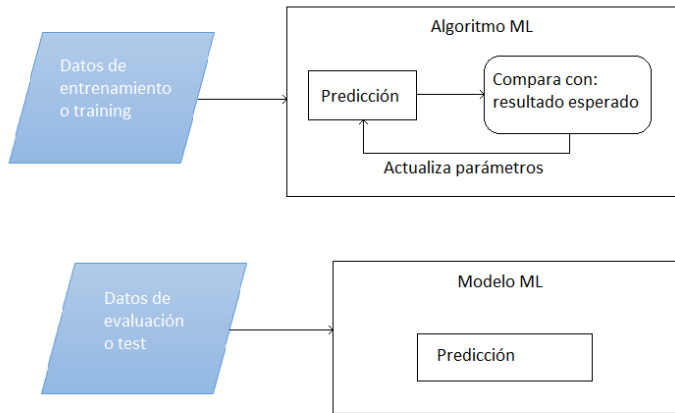


Figura 5.1: Algoritmo de machine learning

Las técnicas de machine learning constan aproximadamente de tres pasos: *entrenamiento o training* en las muestras de simulación para aprender a discriminar entre señal y ruido, *prueba o testing* en muestras de simulación independientes para evaluar los resultados y la evaluación de la salida o predicción en muestras de datos.

5.2. Clasificación en Machine Learning

La toma de decisiones es una parte primordial en estadística aquí la probabilidad juega un papel muy importante permitiendo tomar decisiones óptimas en situaciones que impliquen incertidumbre, como las que se encuentran en identificación de partículas (señal vs ruido).

Supongamos que tenemos un vector de entrada x y un vector t de variables objetivo, se quiere predecir t dado un nuevo valor para x . Si t comprenden variables continuas se tiene un problema de regresión, mientras que si t representa etiquetas de clase se tendrá un problema de clasificación. La distribución de probabilidad conjunta $p(x, t)$ nos brinda la información completa de la incertidumbre asociada a estas variables. Encontrar la probabilidad conjunta $p(x, t)$ a partir de un conjunto de datos

de entrenamiento es un problema de inferencia. Por otra parte, si se está interesado en hacer una predicción específica para el valor de t , esto es precisamente el tema de la teoría de decisión.

Retomando el problema de inferencia consiste en encontrar la distribución conjunta $p(x, C_k)$ para cada clase C_k , o equivalentemente, $p(x, t)$. Esta es una cantidad muy útil e informativa, aunque nos gustaría que esta elección fuera óptima. Es aquí donde entra el tema de la teoría de decisión, nos dice como tomar decisiones óptimas dadas las probabilidades asociadas.

Según Bishop [5], se puede identificar tres enfoques distintos para resolver problemas de decisión. Los cuales se definen a continuación:

1. Modelos probabilísticos:

a) Modelos Generativos:

Primero se determina las densidades condicionales de clase $p(x|C_k)$ para cada clase C_k , siendo $x = (x_1, x_2, \dots, x_N)^T$ individualmente, esto resuelve el problema de inferencia. Por otro lado se infieren por separado las probabilidades de las clases priori $p(C_k)$. Ahora se usa el teorema de Bayes para determinar las probabilidades de la clase posterior $p(C_k|x)$.

$$p(C_k|x) = \frac{p(x|C_k)p(C_k)}{p(x)} \quad (5.2.1)$$

El denominador de este teorema se determina en términos de las cantidades que aparecen en el numerador, de la siguiente forma:

$$p(x) = \sum_{k=1}^K p(x|C_k)p(C_k) \quad (5.2.2)$$

También es posible modelar directamente la distribución conjunta $p(x, C_k)$, luego se normaliza para obtener las probabilidades posteriores y se hace uso de la teoría de decisión para determinar la pertenencia a la clase para cada nueva entrada x . Este enfoque se conoce como modelos probabilísticos generativos, porque a través del muestreo es posible generar datos sintéticos en el espacio de entrada.

Si consideramos el caso de dos clases, la probabilidad posterior para la clase C_1 se puede escribir como:

$$\begin{aligned} p(C_1|x) &= \frac{p(x|C_1)p(C_1)}{p(x|C_1)p(C_1) + p(x|C_2)p(C_2)} \\ &= \frac{1}{1 + \exp(-a)} \\ &= \sigma(a) \end{aligned} \tag{5.2.3}$$

con

$$a = \ln \left(\frac{p(x|C_1)p(C_1)}{p(x|C_2)p(C_2)} \right) \tag{5.2.4}$$

Y $\sigma(a)$ es la llamada función logística sigmoide. El término “sigmoide” significa en forma de S. El sigmoide logístico juega un papel importante en los problemas de clasificación.

b) Modelos Discriminativos:

Un enfoque alternativo al modelo generativo es usar la forma funcional del modelo lineal generalizado explícitamente y determinar sus parámetros directamente usando la máxima verosimilitud. Las soluciones se pueden encontrar por el método de mínimos cuadrados iterativos reponderados.

En este modelo primero se determina las probabilidades de clase posterior $p(C_k|x)$, esto resuelve el problema de inferencia, luego se utiliza la teoría de decisión para asignar cada nueva entrada x a una de las clases C_k .

En este modelo la técnica es modelar directamente $p(C_k|x)$, por ejemplo, representándolos como modelos paramétricos y luego optimizando los parámetros usando un conjunto de entrenamiento.

Una ventaja del modelo discriminativo es que por lo general tendrá menos parámetros adaptativos que determinar. Cuando

los supuestos de densidad condicional de clase dan malas aproximaciones con respecto a las distribuciones reales, este modelo puede conducir a un rendimiento predictivo mejorado.

2. Función Discriminante:

En este modelo se encuentra una función $f(x)$, llamada función discriminante, que asigna directamente a cada entrada x una etiqueta de clase, denominada C_k . Por ejemplo, si se tiene dos clases $C_k = C_1, C_2$, $f(\cdot)$ puede tener un valor binario, de forma que, $f = 0$ representa la clase C_1 y $f = 1$ representa la clase C_2 . En este modelo las probabilidades no juegan ningún papel. Estas funciones solo se encargan de obtener una frontera de decisión.

Para el caso de dos clases se suele utilizar una función discriminante lineal, la cual toma una función lineal del vector de entrada tal que:

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0 \quad (5.2.5)$$

Donde \mathbf{w} se denomina vector de peso y w_0 es un sesgo (diferente al sesgo en el sentido estadístico). Si $y(\mathbf{x}) \geq 0$ el vector de entrada \mathbf{x} se asigna a la clase C_1 , en caso contrario se asigna a la clase C_2 . Así, la relación $y(\mathbf{x}) = 0$ define el límite de decisión que corresponde a un hiperplano de dimensión $(D - 1)$. Si \mathbf{x} es un punto en la superficie de decisión, entonces $y(\mathbf{x}) = 0$, por lo tanto, la distancia normal del origen a la superficie de decisión es la siguiente:

$$\frac{\mathbf{w}^T \mathbf{x}}{\|\mathbf{w}\|} = -\frac{w_0}{\|\mathbf{w}\|} \quad (5.2.6)$$

Así, el parámetro de sesgo w_0 determina la ubicación de la superficie de decisión.

Cada uno de estos tres enfoques tiene sus ventajas y desventajas. Por ejemplo, el enfoque probabilístico generativo, es muy costoso computacionalmente dado que se tiene que encontrar la distribución conjunta sobre x y C_k . Los valores de la clase $p(C_k)$ se puede estimar de los datos del

conjunto de entrenamiento en cada una de las clases. Este enfoque tiene la ventaja que permite encontrar la densidad marginal de los datos $p(x)$, que es deseable para detectar nuevos puntos de datos que tienen baja probabilidad dado el modelo, esto se conoce como la detección de valores atípicos.

Ahora bien, si solo queremos tomar decisiones de clasificación, solo se necesita encontrar las probabilidades posteriores $p(C_k|x)$ y no la distribución conjunta $p(x, C_k)$ este es el enfoque probabilístico discriminativo. El enfoque más sencillo de los tres es el de la función discriminante, tiene la ventaja que combina en un solo problema las etapas de inferencia y decisión. Sin embargo, aquí ya no se tiene acceso a las probabilidades posteriores $p(C_k|x)$, solo nos da el límite de decisión que da la mínima probabilidad de clasificación errónea.

En este estudio, se hará uso de la función discriminante mediante el algoritmo de árboles de decisión. Aunque se entrenaron otros métodos este estudio esta centrado en árboles de decisión.

5.3. Motivación para utilizar métodos multivariantes (MVA)

Tradicionalmente se hacían cortes rectangulares a cada una de las variables y se definía la clasificación como el conjunto de cortes utilizados en todas estas variables, se tiene la desventaja que esto se vuelve ineficiente y mas aún si se tienen muchas variables es más costoso computacionalmente. Se debe comprender bien lo que se está haciendo, porque el rechazo o el corte de eventos de background se está haciendo variable a variable, cuando se hace el corte en una variable cambia toda la distribución y se hace de nuevo el análisis para saber cual es el corte óptimo en las variables restantes, se sigue el mismo procedimiento hasta hacer el corte en la última variable es por ello que es ineficiente. A partir del comportamiento físico de los procesos de señal y de ruido, se pueden postular valores justificados para el corte que disminuyan lo más posible la presencia de eventos de ruido, B, respecto a los sucesos de señal, S.

El objetivo es encontrar una región donde S sea mucho mayor que B, por ello es usual que las cantidades discriminantes sean definidas como

S/B , $S/(S+B)$, $S/\sqrt{S+B}$ entre otras. Entonces una región de corte será mejor que otra cuanto mayor sea el valor de estos discriminantes.

Alternativamente, podemos definir la eficiencia del corte en la señal y el ruido: ϵ_s y ϵ_b . La eficiencia respecto a un proceso se definirá como la proporción de eventos del mismo que cumplen la condición de corte. Un corte será mejor cuanto mayor sea su eficiencia de señal y menor sea su eficiencia de ruido.

El problema es que una sola variable no da valores elevados de significancias, eficiencias y purezas, se realizan varios cortes secuenciales para definir la región de trabajo. A este tipo de análisis que involucran una única variable se dice que son mono-variante o univariante. Estas variables individuales no tienen un gran poder de separación y por ende, sus cortes no son muy óptimos. De hecho, podemos decir que es imposible lograr la significancia, eficiencia y pureza requeridas utilizando cortes en una sola variable y es por este motivo que son necesarios métodos que combinen muchas variables diferentes.

Por lo tanto, es por ello que se utilizan estos métodos multivariados para clasificación, porque combina un conjunto de variables, con buena separación entre señal y ruido en una sola variable que es el resultado que deseamos. Esto permite el rechazo de ruido mediante cortes en una sola "variable." discriminante, generada específicamente para ese propósito. Se pueden hacer varias combinaciones y entrenar varios métodos al mismo tiempo, aparte de eso no deforma el espacio de fase que es algo muy deseable para la física, en general tiene un mejor rendimiento.

Se desarrollaron muchos métodos diferentes con el objetivo de realizar la clasificación de eventos simultáneamente con múltiples variables. Existen varios métodos de MVA entre ellos están Discriminante lineal (LD), Redes Neuronales Artificiales, Boosted Decision Tree (BDT), Maquinas de Vectores de Soporte, Estimador de verosimilitud proyectiva. En el presente trabajo utilizaremos BDT para la clasificación entre señal y ruido, es una de las técnicas más utilizadas para este propósito en física de partículas. Estos métodos se organizaron en un paquete de C++ llamado Toolkit for Multivariate Analysis (TMVA) [3] que se implementa en el marco de ROOT y tiene aplicaciones específicamente en física de altas energías.

El CERN ha desarrollado **ROOT** [36], un entorno de trabajo, orientado a objetos para el análisis de datos a gran escala y que provee métodos estadísticos. Se utilizara ROOT y el paquete TMVA para nuestro análisis. Existen otros paquetes entre ellos están: Neural Bayes y **scikit-learn** que es Machine learning in Python.

5.4. Algoritmos de Machine Learning

En esta sección se describirán algunos algoritmos de clasificación, implementados en ROOT mediante el paquete TMVA. Se dará una descripción del algoritmo utilizado en esta tesis para los resultados de la misma, el cual es BDT. También se dará una breve descripción de dos algoritmos que se comparan con el funcionamiento de BDT.

Estas son algunas referencias de investigaciones que han utilizado aprendizaje supervisado en específico el algoritmo de BDT para la clasificación de señal y ruido: [6], [8], [38], [33], [34], y [28].

5.4.1. Boosted Decision Trees (BDT)

Muchos problemas de clasificación son demasiado complejos para usar cortes lineales en variables de eventos como clasificación y generalmente involucran un ruido muy grande en comparación con la señal.

Un árbol de decisión ([4], [6] y [8]) es una estructura binaria que consta de un nodo raíz y varios nodos de rama y hoja (Observar Figura 5.2). El nodo raíz y cada nodo de rama están asociados con una pregunta binaria que involucra una de las variables discriminatorias. El árbol se atraviesa desde la raíz hasta una de las hojas y el camino se determina por la respuesta a la pregunta asociada con cada nodo posterior. Las hojas representan un evento ya sea como señal o como background.

El procedimiento involucra cortar secuencialmente las variables discriminantes, de esta manera un evento se procesa a través del árbol que conlleva a una respuesta de señal o ruido. Siguiendo este método el espacio de las variables discriminantes se divide en muchas regiones rectangulares, que representan señal o background.

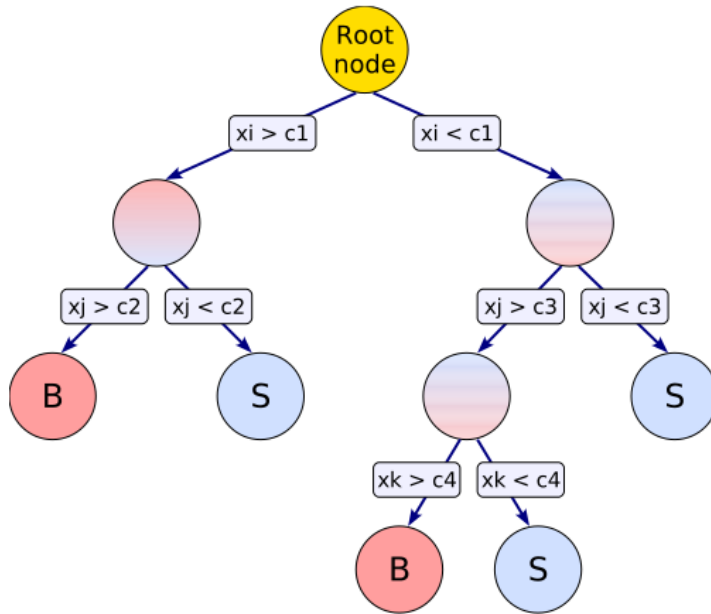


Figura 5.2: Estructura de un árbol de decisión, tomado de [3]

Construcción del BDT

Supongamos que tenemos un conjunto X con propiedades $\{P_1, P_2, \dots, P_n\}$. Sea X el nivel 0 del árbol y consideremos una partición binaria (señal y ruido) de X en función de una de sus propiedades P_i . Así tomamos dos condiciones mutuamente excluyentes entre sí A_1 y A_2 que puede cumplir P_i . Entonces dividimos a X en dos subconjuntos distintos $X_1 = \{x \in X/x_i \in A_1\}$ y $X_2 = \{x \in X/x_i \in A_2\}$. Estos subconjuntos de X forman el nivel 1 del árbol de decisión y se denominan nodos. Siguiendo el mismo proceso, podemos ir generando los siguientes niveles del árbol. Denominamos profundidad al nivel máximo que alcanza el árbol.

Los árboles de decisión se cultivan o crecen respectivamente de manera sucesiva, en la Figura 5.2 se observa la estructura de un árbol comenzando con el nodo raíz, se elige el valor de corte óptimo y la variable discriminante en cada nodo de modo que proporcione la mejor separación entre señal y

background. El entrenamiento de un árbol se detienen en un nodo cuando alcanza un límite mínimo de eventos ($nEventsMin$) en ese nodo. Hojas se les llama a los nodos finales, estas determinan si el evento final es señal o background y se etiqueta de acuerdo con su pureza. Definimos la calidad de la separación como la pureza del nodo con un conjunto de eventos de señal S_i y de ruido B_i así:

$$p_i = \frac{\sum_{x \in S_i} w_x}{\sum_{x \in S_i} w_x + \sum_{x \in B_i} w_x} \quad (5.4.1)$$

Esto no es mas que una forma de expresar $\frac{S}{S+B}$, donde w_x son los pesos asociados a cada evento donde si no viene especificado asumimos que $w_x = 1$. Observamos que P tomará valores cercanos al 0 en caso de que el ruido sea mayoritario y valores próximos a 1 si lo es la señal. Es por ello que si $p_i > 0,5$ el evento se considera señal y si $p_i < 0,5$ se considera background.

La calidad de separación esta definido por la llamada función de impureza algunos ejemplos son los siguientes:

- Índice de Gini: $p_i(1 - p_i) \sum_{x \in N_i} w_x$
- Entropía cruzada: $-p_i \ln(p_i) - (1 - p_i) \ln(1 - p_i)$
- Error de Mis Clasificación: $1 - \max(p_i, 1 - p_i)$
- Significancia estadística: $\frac{S}{\sqrt{S+B}}$

Las tres medidas son similares, pero el índice de Gini y la entropía cruzada son diferenciables y, por lo tanto, más susceptibles de optimización numérica. Además, el índice de Gini y la entropía cruzada son más sensibles al cambio en las probabilidades de los nodos que el error de clasificación errónea. El índice de Gini y la entropía cruzada son similares. El que usaremos en nuestro análisis es el índice de Gini.

Boosting

Una deficiencia de los BDT es su inestabilidad con respecto a las fluctuaciones estadísticas en la muestra de entrenamiento de la que se deriva la estructura del árbol. Por ejemplo, si dos variables de entrada exhiben un

poder de separación similar, una fluctuación en la muestra de entrenamiento puede hacer que el algoritmo de crecimiento del árbol decida dividirse en una variable, mientras que la otra pudo haberse seleccionado sin esa fluctuación. En tal caso, toda la estructura del árbol se modifica debajo de este nodo, lo que posiblemente también resulte en una respuesta de clasificador sustancialmente diferente.

Este problema se supera construyendo un bosque de árboles de decisión y clasificando un evento por mayoría de votos de las clasificaciones realizadas por cada árbol del bosque. Todos los árboles del bosque se derivan de la misma muestra de formación y los eventos se someten posteriormente al llamado Boosting (refuerzo) un procedimiento que modifica sus pesos en la muestra.

El Boosting aumenta la estabilidad estadística del clasificador, hace que los árboles sean mas robustos y puede mejorar drásticamente el rendimiento de la separación en comparación con un único árbol de decisión. Sin embargo, se pierde la ventaja de la interpretación sencilla del árbol de decisión. En muchas ocasiones el boosting funciona mejor si se aplica a árboles (clasificadores) que tomados individualmente ya que no tienen mucho poder de clasificación, estos llamados “clasificadores débiles” son árboles pequeños de crecimiento limitado a una profundidad de tan solo 2 niveles. Al limitar la profundidad del árbol durante el proceso de construcción del árbol (entrenamiento) se elimina casi por completo la tendencia al sobre-entrenamiento para árboles de decisión simple que generalmente se cultivan a una gran profundidad y luego se podan. Para mayor información del Boosting consultar el capítulo 10 de [37]

TMVA proporciona tres tipos de boosting diferentes:

- Adaptive Boosting(AdaBoost)
- Gradient Boosting
- Bagging Boosting

El boosting utilizado en el BDT es el conocido como AdaBoost, el cual se describe a continuación, basado en la referencia [4].

Si tenemos N eventos totales en la muestra, suponga que construimos k árboles, T_k . Definimos el error del árbol k así:

$$err_k = \frac{\sum_{i=1}^N w_i I(y_i \neq T_k(x_i))}{\sum_{i=1}^N w_i} \quad (5.4.2)$$

donde,

- w_i : El peso del i -ésimo evento
- x_i : Conjunto de variables discriminantes para el i -ésimo evento
- $y_i = 1$ si el i -ésimo evento es un evento de señal y $y_i = -1$ si el i -ésimo evento es un evento de ruido.
- $T_k(x_i) = 1$ si el evento se encuentra en una hoja tipo señal en el árbol i -ésimo y $T_k(x_i) = -1$ si el evento se encuentra en una hoja tipo ruido en el árbol i -ésimo.
- $I(y_i \neq T_k(x_i)) = 1$ si $y_i \neq T_k(x_i)$ y 0 si $y_i = T_k(x_i)$, es decir es igual a 1 si la hoja en la que se encuentra el evento es del mismo tipo que este y 0 en otro caso

Lo que mide err_k es la proporción de eventos mal clasificados. Mientras mayor sea err_k , el árbol obtenido disminuirá su calidad. Ahora definamos el peso del árbol como:

$$\alpha_k = \beta \ln \left(\frac{1 - err_k}{err_k} \right) \quad (5.4.3)$$

Donde β representa la intensidad del boost y es un parámetro adicional del BDT, $\beta = 1$ es el valor utilizado en el método AdaBoost estándar. Por último, cada uno de los pesos del evento se cambia de la siguiente forma:

$$w_i \rightarrow w_i \exp^{\alpha_k I(y_i \neq T_k(x_i))} \quad (5.4.4)$$

Y se normaliza $w_i \rightarrow \frac{w_i}{\sum_{i=1}^{N_T} w_i}$ De forma que los eventos mal identificados ven su peso aumentado y los bien identificados lo ven disminuido. Adicionalmente, cuanto mejor sea el árbol (mayor valor de α), más se modifican los pesos, con lo que se consigue variar más sensiblemente los siguientes resultados para diversificar el poder de separación del BDT.

Con todas estas definiciones estamos en condiciones de definir la variable discriminante que utiliza los datos de todos los árboles generados. Para cada uno de los eventos estudiados se define la variable BDT o respuesta del BDT:

$$BDT(x) = \sum_{k=1}^{N_T} \alpha_k T_k(x) \quad (5.4.5)$$

Donde N_T es el total de arboles considerados.

Overtraining

Cuando los clasificadores están demasiado adaptados a una muestra de entrenamiento específico, muestras similares pueden tener una respuesta diferente a la del clasificador entrenado. Cuando esto sucede estamos hablando de sobre entrenamiento es problemático ya que los datos reales se clasificaran de manera diferente a la muestra de entrenamiento y a la distribución de salida del BDT no se puede utilizar para determinar el valor de corte óptimo.

Para nuestro análisis tenemos muestras adicionales para verificar la validez del BDT, puesto que no se pueden utilizar las mismas muestras con las que fue entrenado dado que se encuentra optimizado con las mismas. Por ello es que utilizamos muestras independientes a las que fueron utilizadas en el entrenamiento y sus resultados deben ser similares a los que se obtienen al aplicarlo con las muestras de entrenamiento. Esta similitud se puede comprobar con pruebas de hipótesis sobre la distribución la variable de BDT tanto para test como para entrenamiento, pruebas tales como χ^2 y kolmogorov-Smirnov de doble distribución. Es indispensable evitar la presencia de overtraining para obtener arboles que sean buenos discrimi-

nantes.

Debemos tener en cuenta también la aparición de variables correlacionadas, cuando dos variables están correlacionadas decimos que sus distribuciones son dependientes. En BDT si dos variables se encuentran fuertemente correlacionadas el valor de una determina, prácticamente, el valor de la otra, es decir, una nos da la misma información que la otra y por tanto no puede proporcionar un poder de separación mayor al propio BDT. Lo ideal es eliminar una de estas variables correlacionadas lo cual implica una aceleración en el proceso de entrenamiento.

5.4.2. Análisis Discriminante lineal (LD)

Un discriminante es una función que toma un vector de entrada x y lo asigna a una de las k clases, llamadas C_k . Estos discriminantes son lineales, es decir, las superficies de decisión son hiperplanos. Este análisis realiza la clasificación mediante un modelo lineal, donde la función discriminante ($y(x)$) es lineal en los parámetros β

$$y(x) = x^T \beta + \beta_0 \tag{5.4.6}$$

Donde β_0 es un sesgo (que no debe confundirse con el sesgo en el sentido estadístico). A veces se denomina umbral al negativo del sesgo. Así un vector de entrada x se asigna a la clase C_1 (señal) si $y(x) \leq 0$ y a la clase C_2 (ruido) si $y(x) < 0$. Este método es equivalente al discriminante de Fisher, que busca maximizar la varianza entre clases y la varianza dentro de las clases de los datos proyectados en una sola dimensión.

Supongamos que hay $m + 1$ parámetros $\beta_0, \beta_1, \dots, \beta_m$ utilizados para estimar un conjunto de entrenamiento compuesto por n eventos, se define β así:

$$Y = X\beta \tag{5.4.7}$$

Visto matricialmente (β_0 es absorbido en el vector β)

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1m} \\ 1 & x_{21} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nm} \end{pmatrix} \quad (5.4.8)$$

donde, la matriz Y es tal que $y_i = 1$ si el i -ésimo evento pertenece a la clase de señal y $y_i = 0$ si el i -ésimo evento pertenece a la clase de ruido. Ahora bien, se aplica el método de mínimos cuadrados y se obtienen las ecuaciones normales para el problema de clasificación. Se multiplica la matriz X^T a ambos lados de la igualdad de la ecuación 5.4.7

$$Y = X\beta \iff X^T Y = X^T X\beta \iff \beta = (X^T X)^{-1} X^T Y \quad (5.4.9)$$

La matriz $(X^T X)^{-1} X^T$ se le conoce como la pseudo inversa de Moore-Penrose de X , con la condición de que X tenga rango completo. En estos algoritmos siempre se incluyen pesos a los datos, por lo tanto, se incluye una matriz de peso diagonal W que dando las ecuaciones de la siguiente manera

$$\beta = (X^T W X)^{-1} X^T W Y \quad (5.4.10)$$

El LD puede ser interpretado geoméricamente determinando la frontera de decisión lineal, la recta que maximiza la separabilidad de las 2 clases, consideremos x_1 y x_2 en la frontera de decisión, así $y(x_1) = y(x_2) = 0$, por ende $(x_1 - x_2)^T \beta = 0$, donde β es un vector ortogonal.

5.4.3. Estimación de Verosimilitud Proyectiva

Este método se basa en construir un modelo a partir de funciones de densidad de probabilidad (PDF) que reproduzca las variables de entrada para señal y ruido. La verosimilitud de que un evento sea señal se obtiene multiplicando las PDF de todas las variables de entrada, las cuales son independientes. Este enfoque suele llamarse “Estimador ingenuo de Bayes”, debido a que no toma en cuenta las correlaciones entre las variables. La razón de verosimilitud $y_L(i)$ para cada evento i esta definida como:

$$y_L(i) = \frac{\prod_{k=1}^{n_{var}} p_k^{signal}(x_k(i))}{\sum_U \left(\prod_{k=1}^{n_{var}} p_k^U(x_k(i)) \right)} \quad (5.4.11)$$

donde, $p_k^{signal}(x_k(i))$ es la distribución de probabilidad de señal para la k -ésima variable de entrada x_k , $U = \{señal, ruido\}$. Ahora bien, normalizando las distribuciones de probabilidad tenemos que:

$$\int_{-\infty}^{+\infty} \sum_U p_k^U(x_k) dx_k \quad (5.4.12)$$

Debido que, por lo general la forma paramétrica de las PDF se desconoce, estas se aproximan a partir de los datos de entrenamiento mediante funciones no paramétricas, que se eligen para cada variable y son splines polinómicos de varios grados ajustados a histogramas.

Para mayor información de los métodos y su utilización en TMVA ver [3].

5.5. TMVA: Evaluación del desempeño de la clasificación

Al entrenar una función análisis multivariante (MVA) con el paquete de TMVA [3], los datos de entrada deben discriminar claramente entre la señal y los eventos de ruido, para enseñar correctamente al algoritmo qué tipo de eventos necesita seleccionar.

Como fue mencionado en la sección 5.3 TMVA entrena distintos clasificadores al mismo tiempo. Todos estos clasificadores ven los mismos datos de entrenamiento y luego se prueban todos con los mismos datos de prueba, que es un conjunto de datos independiente de los datos de entrenamiento como se mencionó en la sección 5.1.

Después del entrenamiento, los clasificadores se someten a pruebas y evaluación para medir su desempeño. Es difícil decidir cual es el clasificador óptimo que se utilizará para un análisis determinado. Para ayudarnos

hacer esta elección, TMVA calcula y muestra una serie de cantidades de referencia para comparar los clasificadores utilizando la muestra de prueba independiente. Estas cantidades son:

- La eficiencia de la señal en tres eficiencias de ruido representativas (la eficiencia es igual a 1 - rechazo) obtenidas de un corte en la salida del clasificador. También se proporciona el área del rechazo de ruido frente a la función de eficiencia de la señal (cuanto mayor sea el área, mejor será el rendimiento general), estas son las denominadas curvas ROC.
- La separación $\langle S^2 \rangle$ de una variable dada, esta definida por la integral

$$\langle S^2 \rangle = \frac{1}{2} \int \frac{(\hat{y}_S(y) - \hat{y}_B(y))^2}{\hat{y}_S(y) + \hat{y}_B(y)} dy \quad (5.5.1)$$

Donde \hat{y}_S y \hat{y}_B son las funciones de densidad de probabilidad (PDF) de señal y de ruido para una variable y dada. La separación es cero para señales y ruido idénticos, y es uno para formas sin superposición.

5.5.1. Curvas ROC

La curva ROC que significa característica operativa del receptor (término del procesamiento de señales), traza la tasa de falsos negativos frente a la tasa de verdaderos positivos de un sistema para un clasificador binario.

El lema de Neyman - Pearson definido en la sección de pruebas de hipótesis 2.3.2, define la razón de verosimilitud como criterio de selección, da para cada eficiencia de selección el mejor rechazo de ruido. En el Figura 5.3 se observa un ejemplo de las curvas ROC generadas por TMVA, este gráfico muestra normalmente el rechazo de ruido frente a la eficiencia de la señal, en el cual cada línea representa un clasificador (fisher, MLP, BDT, PDERS, Likelihood). En la figura 5.4 se observa que un clasificador es mejor cuanto mayor sea su área de integración bajo la curva, es decir el clasificador óptimo maximiza el área bajo la curva.

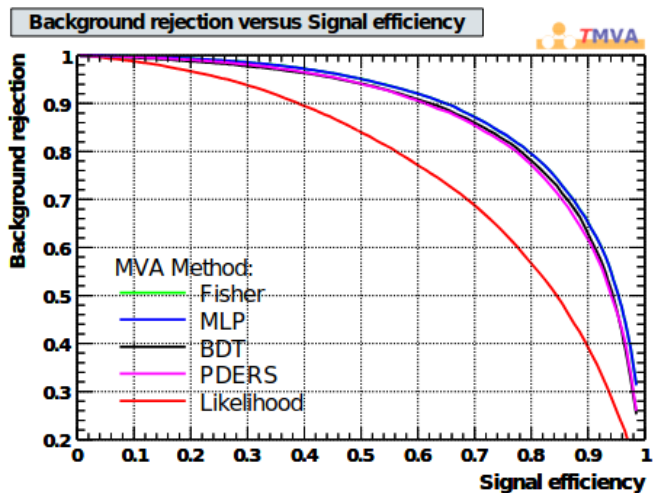


Figura 5.3: Ejemplo del rechazo de ruido vs eficiencia de la señal (“curva ROC”) obtenida cortando las salidas del clasificador para los eventos de la muestra de prueba. Imagen tomada de [3]

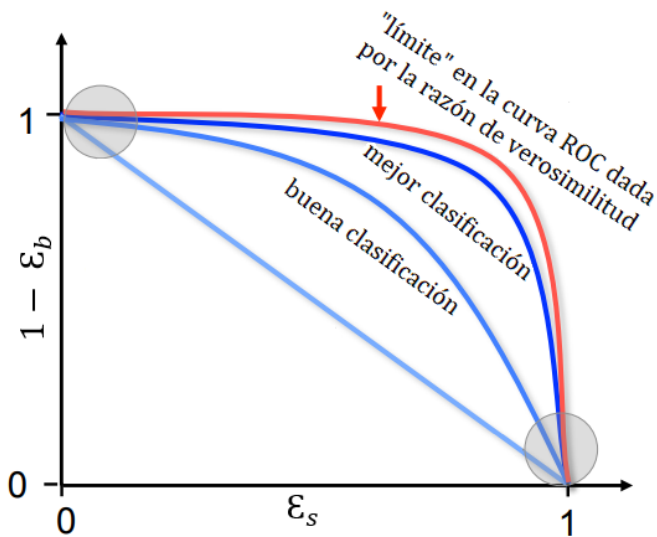


Figura 5.4: Determinación del mejor clasificador mediante las áreas bajo la curva ROC

Capítulo 6

Análisis y Resultados

Estos son los resultados obtenidos en el canal $B \rightarrow \pi^+\pi^+\pi^-$ y $B \rightarrow K^+K^-\pi^+$. En la sección 6.1 se detalla la estrategia de análisis. En la sección 6.2 se describe la selección de análisis multivariante, seguidamente, en las secciones 6.2.1, 6.2.2 y 6.2.3 es descrito el entrenamiento, optimización y aplicación del corte óptimo de BDT, respectivamente. Posteriormente en la sección 6.3 es presentada la selección de identificación de partículas. Finalizando en las secciones 6.3.1 y 6.3.2 que detallan la optimización y la aplicación del criterio PID.

6.1. Estrategia de análisis

Se utilizó la estrategia de análisis descrita en la Figura 6.1, donde se detalla el software de análisis. Luego la selección de análisis multivariante para la reducción del ruido combinatorial donde se utilizó algoritmos de machine learning basados en Boosted Decision Tree, se hizo un proceso de optimización para mis-identificación de partículas donde se construyen figuras de méritos. Finalmente se realiza la aplicación de los cortes de BDT y PID a la distribución de probabilidad de masa invariante.

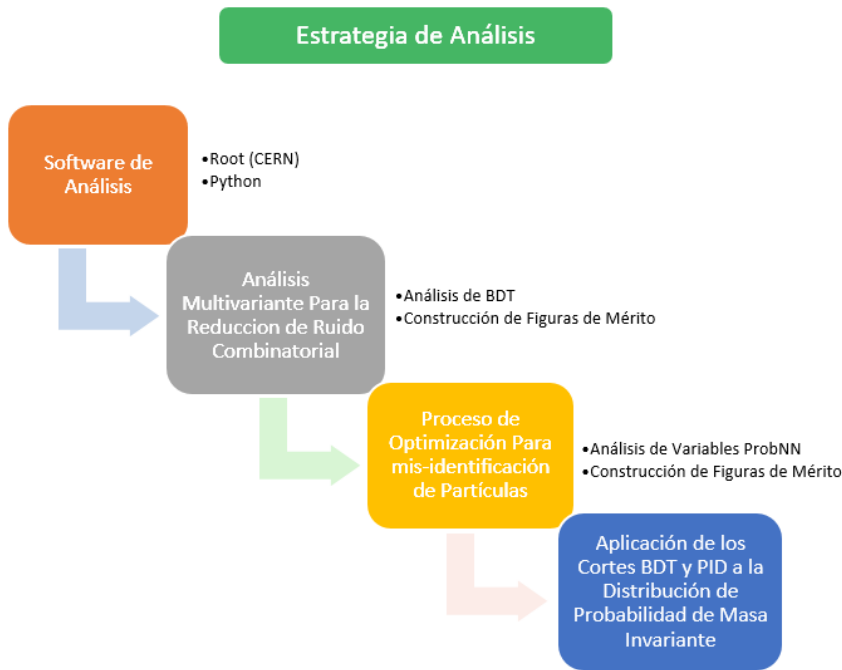


Figura 6.1: Estrategia de análisis

6.2. Selección Análisis Multivariante

Las distintas técnicas de análisis multivariante se basan en algoritmos de aprendizaje supervisados, que utilizan datos de entrenamiento con una salida conocida con el objetivo de producir un modelo que sea capaz de generalizar a partir de los datos de entrenamiento a nuevas situaciones. Pueden encontrar más información sobre MVA en la sección 5.2.

6.2.1. Entrenamiento de BDT

Las muestras de datos y muestras de simulación utilizadas para el entrenamiento de BDT, están descritos en la sección 4.3. Teniendo en cuenta para este entrenamiento se utilizó el lado derecho de la masa invariante,

región de alta masa ($B_m > 5400$)

Se seleccionó las mejores variables discriminantes que dan el mayor poder de separación entre señal y ruido, descritas en el apéndice A.1, que se utilizan como variables de entrada en el entrenamiento de BDT para producir una función, que por si sola tiene mayor poder de separación que todas las variables utilizadas para entrenarla. El objetivo es realizar el corte en una sola variable generada con ese propósito que permita el rechazo de ruido. Se utilizó el paquete de TMVA implementado en ROOT que esta diseñado para aplicaciones de física de altas energías, desarrollado en el CERN. Los gráficos generados por el TMVA que muestran las distribuciones de señal y ruido de las variables de entrada se pueden ver en la Figura 6.2 para el canal $B^+ \rightarrow \pi^+\pi^+\pi^-$ y 6.3 para el canal $B^+ \rightarrow K^+K^-\pi^+$.

A continuación, en la Figuras 6.4 y 6.5 se muestra las matrices de correlación para ruido y señal respectivamente del canal $B^+ \rightarrow \pi^+\pi^+\pi^-$, también en las Figuras 6.6 y 6.7 se muestra las matrices de correlación para ruido y señal respectivamente del canal $B^+ \rightarrow K^+K^-\pi^+$, todas generadas por TMVA para el entrenamiento de 2015 y 2016. En el canal $B \rightarrow \pi^+\pi^+\pi^-$ se visualiza que solamente hay dos variables que tienen una correlación moderada de 60 % las cuales son d_3 y B_P , en general la correlación es baja entre las variables lo cual es deseable. Obtenemos resultados similares al visualizar el canal $B^+ \rightarrow K^+K^-\pi^+$.

Hay muchas técnicas para producir el entrenamiento de análisis multivariante (MVA), para este análisis se comparó el desempeño de 7 diferentes técnicas, las cuales son: Boosted decision tree (BDT), Multilayer Perceptron (MLP), Linear Discriminants (LD), likelihood y BDT con algunas modificaciones por ejemplo BDTPCA (BDT con variables transformadas en componentes principales) y BDTG (BDT con el boosting del gradiente).

Se ejecuto el entrenamiento de MVAs con el paquete TMVA con las variables que son mejores discriminantes entre señal y ruido. Para los datos de training (entrenamiento) tomamos el 70 % y para los datos de test (prueba) se tomó el 30 %. Es importante resaltar que los datos que utilizamos en el entrenamiento no pueden ser utilizados para el análisis por esta razón solo tomamos una porción de los datos reales para el entrenamiento.

CAPÍTULO 6. ANÁLISIS Y RESULTADOS

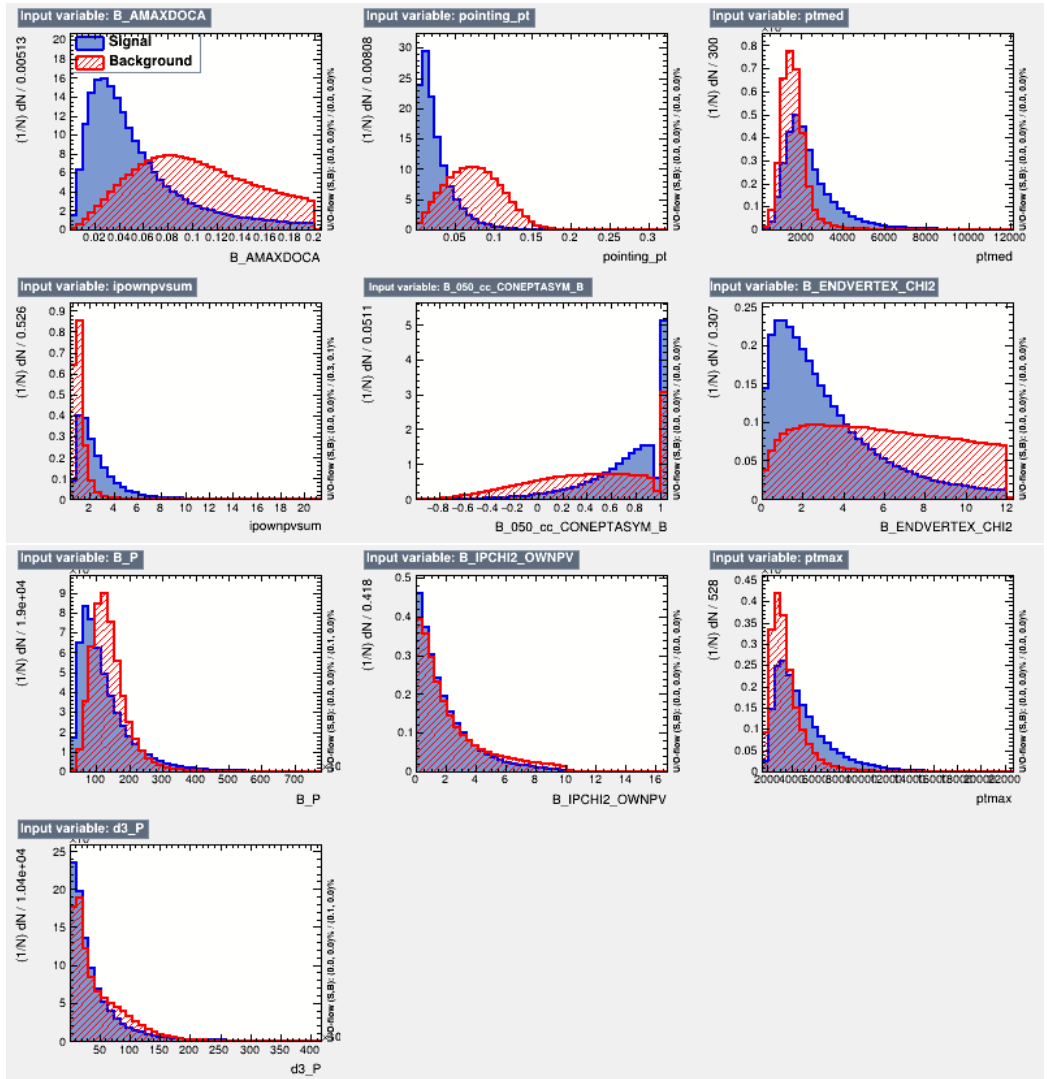


Figura 6.2: Distribuciones de probabilidad (PDFs) de las variables de entrada para el entrenamiento del análisis multivariante, para el canal $B^+ \rightarrow \pi^+ \pi^+ \pi^-$. La señal se representa en azul y el ruido en rojo.

CAPÍTULO 6. ANÁLISIS Y RESULTADOS

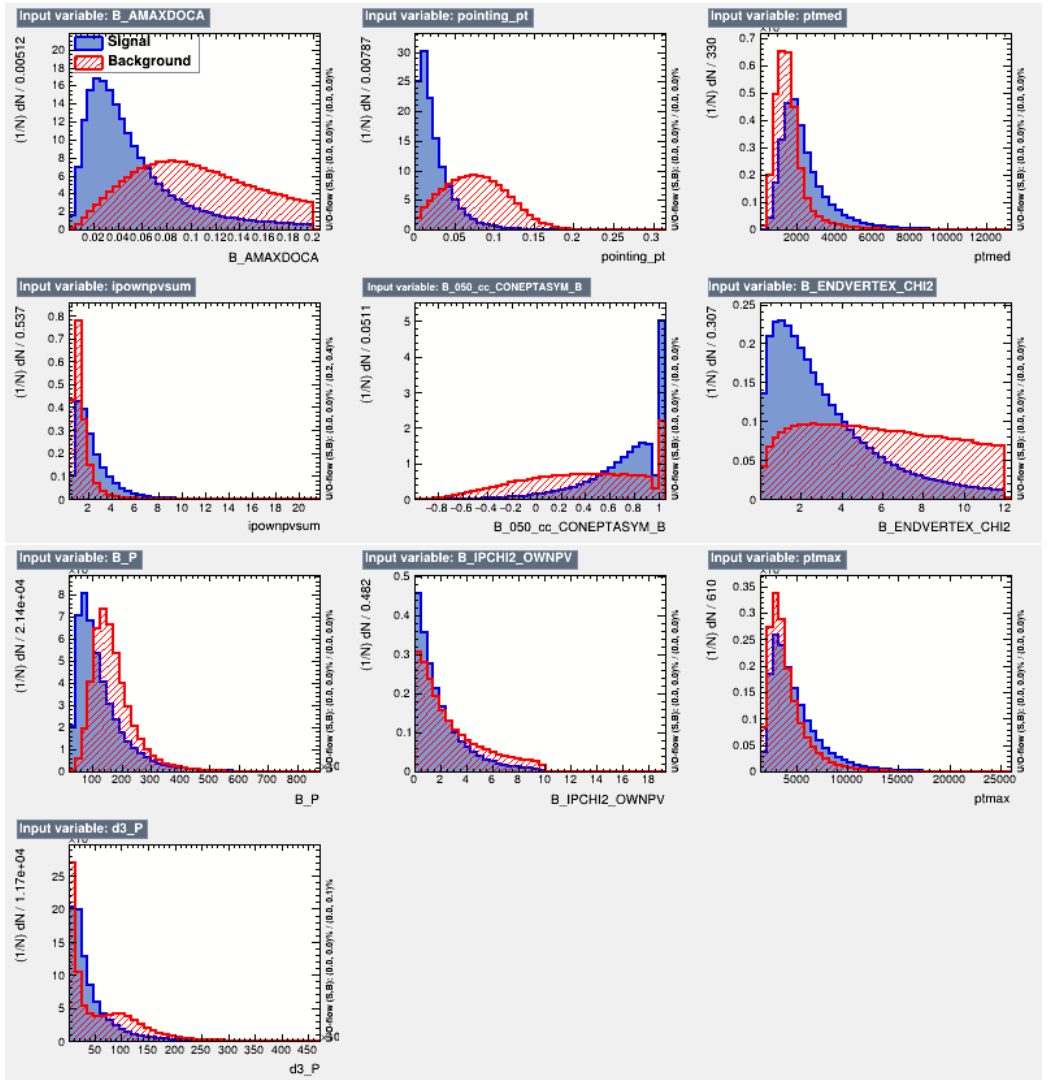


Figura 6.3: Distribuciones de probabilidad (PDFs) de las variables de entrada para el entrenamiento del análisis multivariante, para el canal $B^+ \rightarrow K^+K^-\pi^+$. La señal se representa en azul y el ruido en rojo.

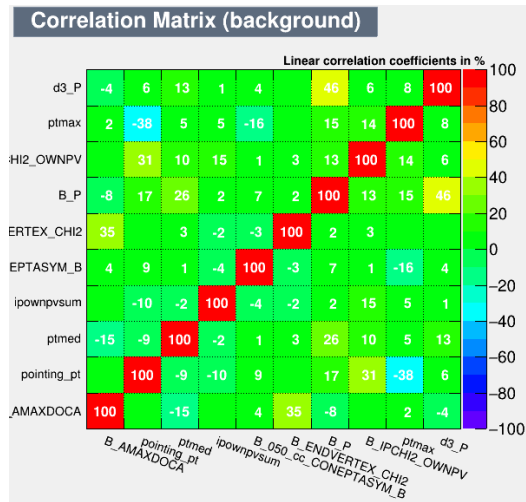


Figura 6.4: Matriz de correlación de ruido para el canal $B^+ \rightarrow \pi^+\pi^+\pi^-$ de las 10 variables seleccionadas para el entrenamiento.

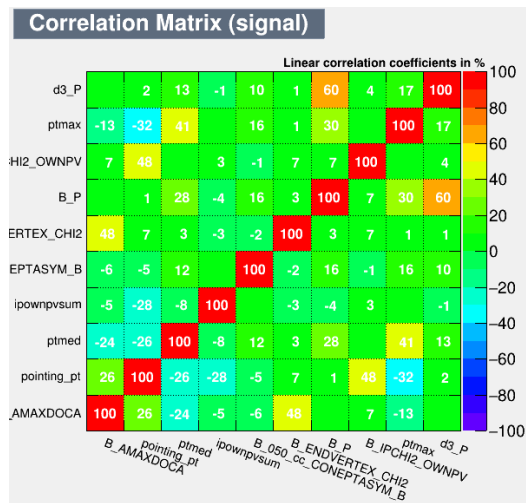


Figura 6.5: Matriz de correlación de señal para el canal $B^+ \rightarrow \pi^+\pi^+\pi^-$ de las 10 variables seleccionadas para el entrenamiento.

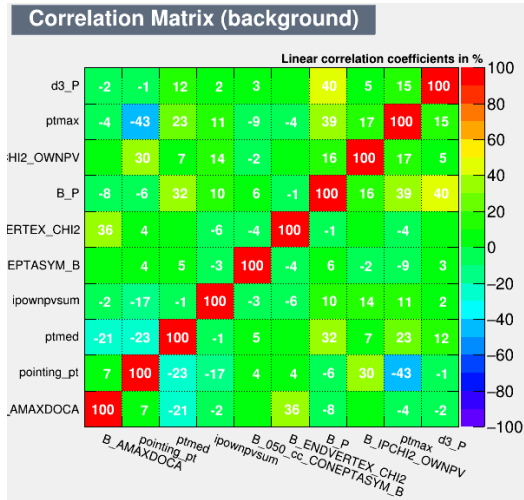


Figura 6.6: Matriz de correlación de ruido para el canal $B^+ \rightarrow K^+K^-\pi^+$ de las 10 variables seleccionadas para el entrenamiento.

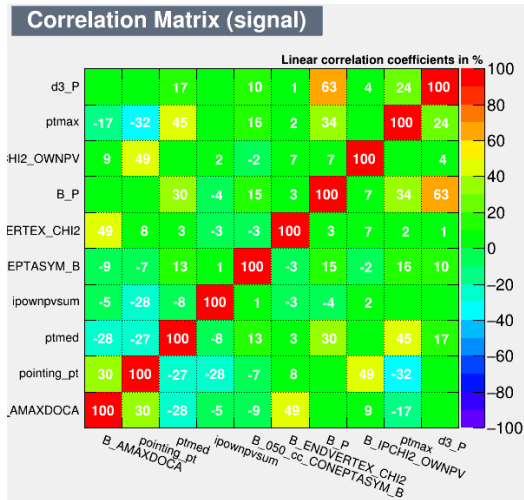


Figura 6.7: Matriz de correlación de señal para el canal $B^+ \rightarrow K^+K^-\pi^+$ de las 10 variables seleccionadas para el entrenamiento.

Las curvas ROC generadas por TMVA nos muestran la eficiencia de la señal frente al rechazo de ruido (ver secciones 2.3.3 y 5.5.1), para cada técnica de MVA utilizada se puede ver en la Figuras 6.8 y 6.9 para el canal $B^+ \rightarrow \pi^+\pi^+\pi^-$ y $B^+ \rightarrow K^+K^-\pi^+$ respectivamente. La técnica de MVA seleccionada debe maximizar tanto la eficiencia de la señal como el rechazo de ruido, el valor de las curvas ROC esta entre 0 y 1 entre mas cercano a 1 este es mejor.

Para ambos canales $B^+ \rightarrow \pi^+\pi^+\pi^-$ y $B^+ \rightarrow K^+K^-\pi^+$, vemos que el BDT es notablemente mejor que las demás técnicas utilizadas, teniendo un comportamiento similar al BDTPCA5100, BDTG y BDTPCA. Seleccionamos el BDT como mejor opción. Cuando comparamos el BDT con MLP, están totalmente superpuestos, pero el BDT tiene un mejor rendimiento con menos tiempo de calculo para el entrenamiento. Se realizarón verificaciones después del entrenamiento para asegurarse de que no hubo sesgo debido al sobreentrenamiento, es mostrado en el apéndice A.2.

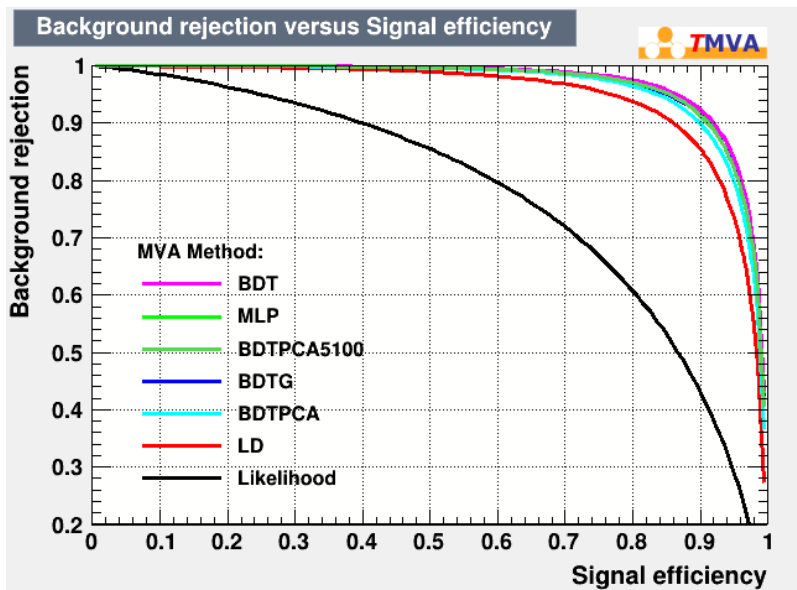


Figura 6.8: Curvas ROC-canal $B^+ \rightarrow \pi^+\pi^+\pi^-$: Eficiencia de la Señal frente al Rechazo de Ruido.

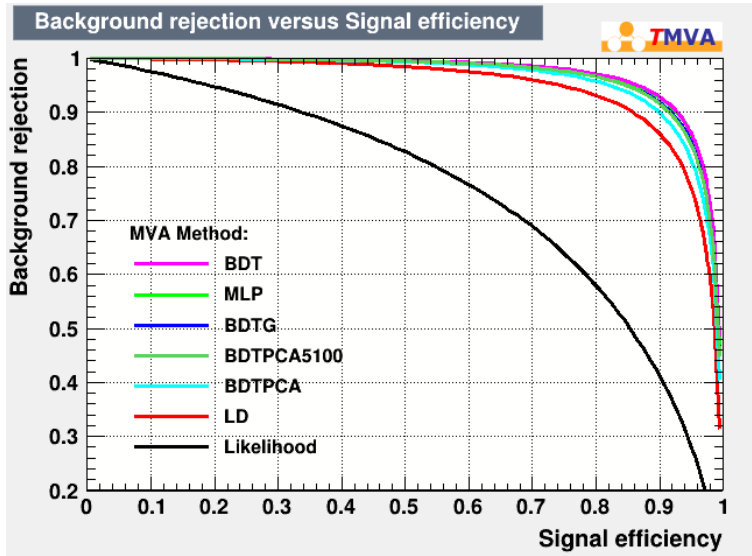


Figura 6.9: Curvas ROC- $B^+ \rightarrow K^+K^-\pi^+$: Eficiencia de la Señal frente al Rechazo de Ruido.

6.2.2. Optimización de BDT

El corte óptimo se determina mediante un estudio de optimización descrito en la sección 4.6. Comenzaremos observando en la Figura 6.10 y 6.11 la distribución de probabilidad de la salida de BDT para el canal $B^+ \rightarrow \pi^+\pi^+\pi^-$ y el canal $B^+ \rightarrow K^+K^-\pi^+$ respectivamente, donde se aprecia que el corte óptimo puede estar alrededor de cero.

El objetivo es buscar la máxima eficiencia de la señal, es decir conservar el mayor número posible de eventos de señal y reducir los eventos de ruido. En la Figura 6.12 y 6.13 generada por TMVA se muestra las eficiencias en función del corte aplicado en la salida de BDT para el canal $B^+ \rightarrow \pi^+\pi^+\pi^-$ y $B^+ \rightarrow K^+K^-\pi^+$ respectivamente. Este gráfico nos da una idea del corte donde se alcanza la máxima significancia estadística (curva verde) y vemos que esta nuevamente alrededor de cero para ambos canales.

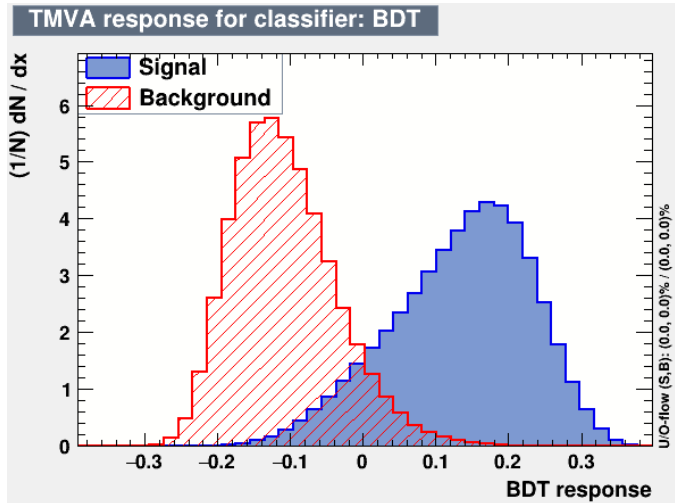


Figura 6.10: Distribución de probabilidad de BDT - $B^+ \rightarrow \pi^+ \pi^+ \pi^-$ año 2015 y 2016.

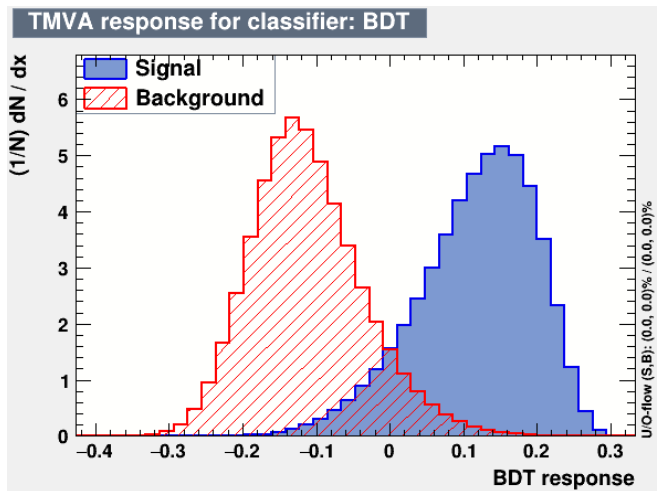


Figura 6.11: Distribución de probabilidad de BDT - $B^+ \rightarrow K^+ K^- \pi^+$ año 2015 y 2016.

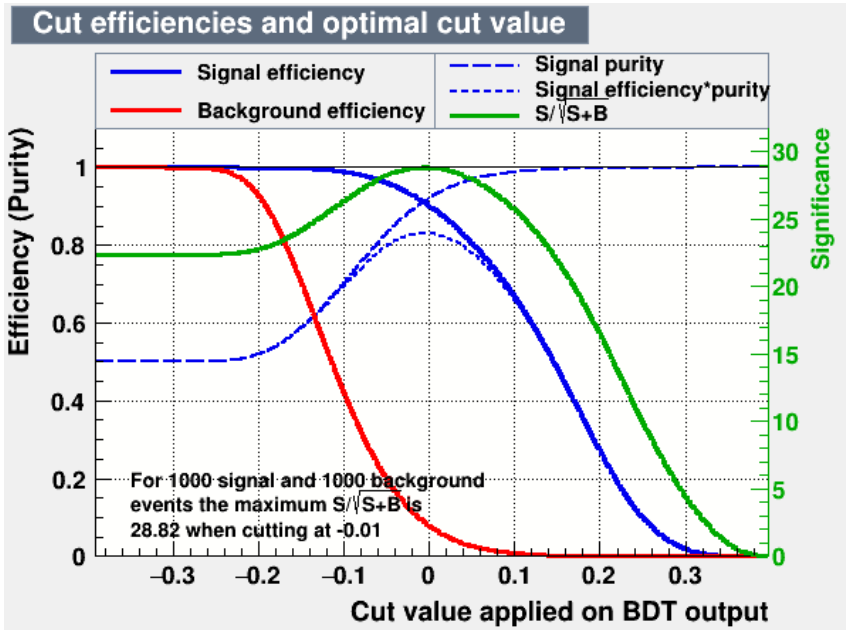


Figura 6.12: Eficiencias en Función del Corte aplicado en la Salida de BDT, canal $B^+ \rightarrow \pi^+\pi^+\pi^-$

Sin embargo buscamos el corte óptimo que nos ayude a reducir al máximo el numero de eventos de ruido conservando la máxima eficiencia de la señal. Esto puede aumentar la pureza de los datos aceptados con una baja pérdida de eficiencia. Calculando los valores de la señal y el ruido para varios puntos de corte, se puede determinar el corte óptimo que nos de la mayor significancia estadística.

Primero se encontrará el corte óptimo para el canal $B^+ \rightarrow \pi^+\pi^+\pi^-$. La Tabla 6.1 nos muestra algunos cortes de los posibles dentro del intervalo $[-0.2, 0.2]$ y sus valores de pureza, eficiencia y significancia estadística definidas en la sección de optimización 4.6.

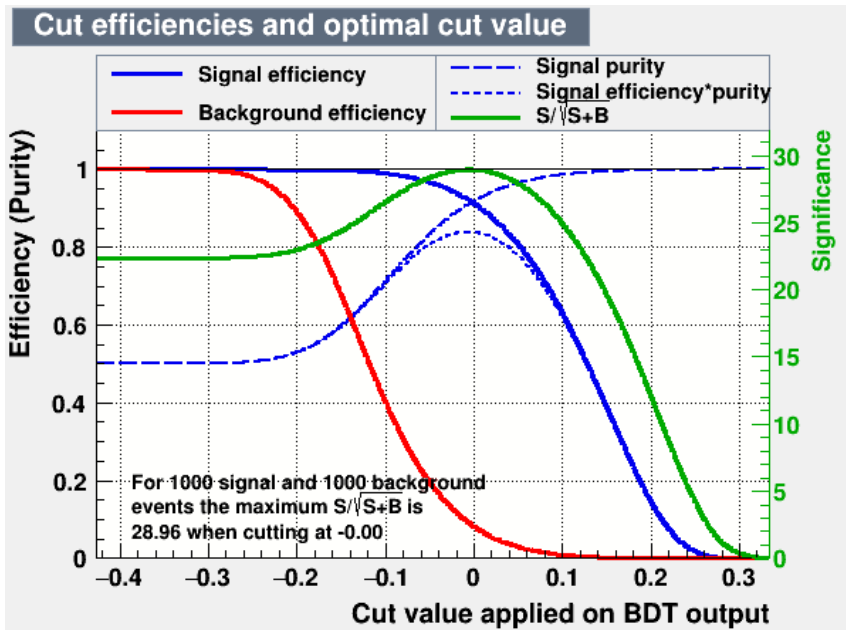


Figura 6.13: Eficiencias en Función del Corte aplicado en la Salida de BDT, $B^+ \rightarrow K^+ K^- \pi^+$

Corte BDT	#Señal+Ruido	#Señal	Eficiencia	Pureza	Significancia
-0.2	327904	2269978	0.9999	0.8738	3964.13
-0.1	191095	2261669	0.9963	0.9187	5173.74
0.044	76091	1999730	0.8809	0.8486	7249.44
0.1	58106	1669652	0.7355	0.7108	6926.53
0.2	21914	700901	0.3088	0.2994	4734.74

Tabla 6.1: Algunos de los posibles cortes para la salida de BDT con sus atributos para el canal $B^+ \rightarrow \pi^+ \pi^+ \pi^-$

Podemos observar que la mayor significancia de los datos se obtiene en $BDT > 0,044$ esto se visualiza en la figura de mérito (FoM) de la significancia estadística (ver Figura 6.14), logrando una eficiencia de la señal del 88.09% (ver Figura 6.16 (a)). En el corte de $BDT > 0,044$ se logra

alcanzar una pureza del 84.8% misma que se visualiza en la Figura 6.16 (b).

En la Figura 6.15 podemos observar nuevamente la curva ROC solamente de BDT, nos brinda la eficiencia de la señal frente al rechazo de background. Para un 88.09% de la eficiencia de la señal se logra reducir en un 97% los eventos de ruido, lo cual es un excelente resultado para los análisis posteriores.

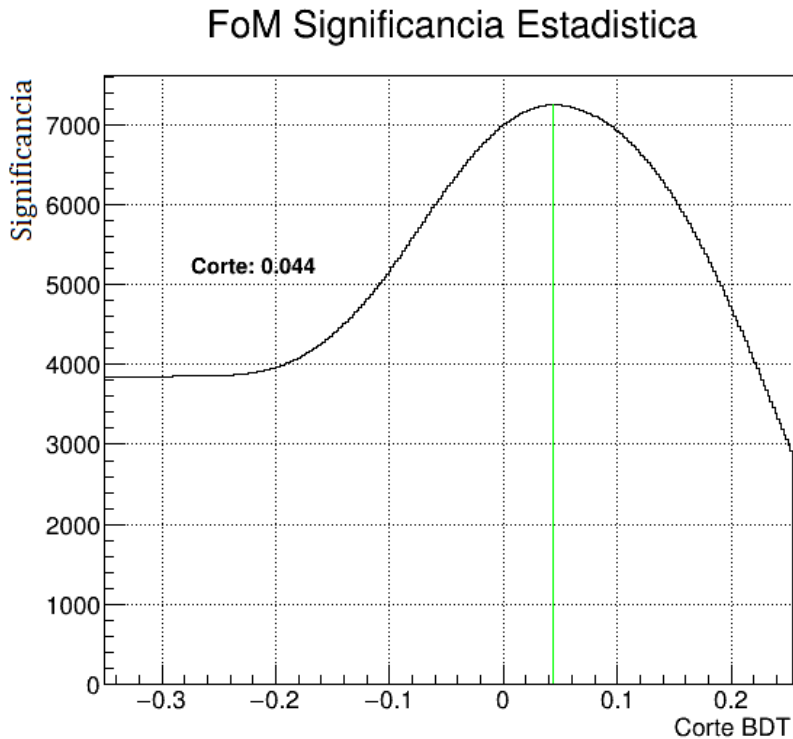


Figura 6.14: Figura de Merito (FoM) de la significancia estadística canal $B^+ \rightarrow \pi^+\pi^+\pi^-$

Ahora se encontrará el corte óptimo para el canal $B^+ \rightarrow K^+K^-\pi^+$. La Tabla 6.2 nos muestra algunos cortes de los posibles dentro del intervalo $[-0.2, 0.2]$ y sus valores de pureza, eficiencia y significancia estadística.

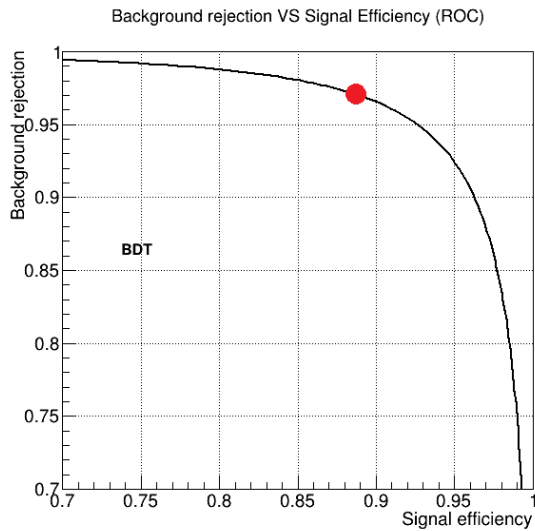


Figura 6.15: Eficiencia de la señal frente al rechazo de background canal $B^+ \rightarrow \pi^+\pi^+\pi^-$

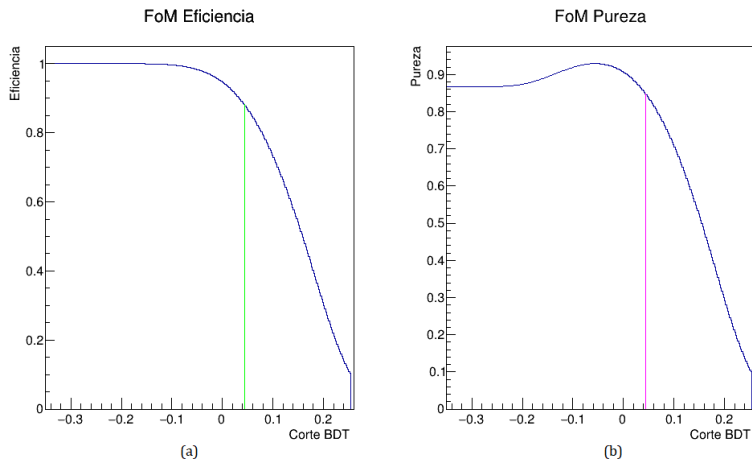


Figura 6.16: En la figura (a): FoM de la eficiencia. (b): FoM pureza

Corte BDT	#Señal+Ruido	#Señal	Eficiencia	Pureza	Significancia
-0.2	528324	3109000	0.9999	0.8547	4277.31
-0.1	338960	3099655	0.9969	0.8987	5324.01
0.06	134068	2576139	0.8286	0.7876	7035.69
0.1	91810	1093760	0.6734	0.6451	6910.06
0.2	11710	465457	0.1497	0.1460	4301.31

Tabla 6.2: Algunos de los posibles cortes para la salida de BDT con sus atributos para el canal $B \rightarrow K^+ K^- \pi^+$

Podemos observar que la mayor significancia de los datos se obtiene en $BDT > 0,06$ esto se visualiza en la figura de mérito (FoM) de la significancia estadística (ver Figura 6.17), logrando una eficiencia de la señal del 82.86 %.

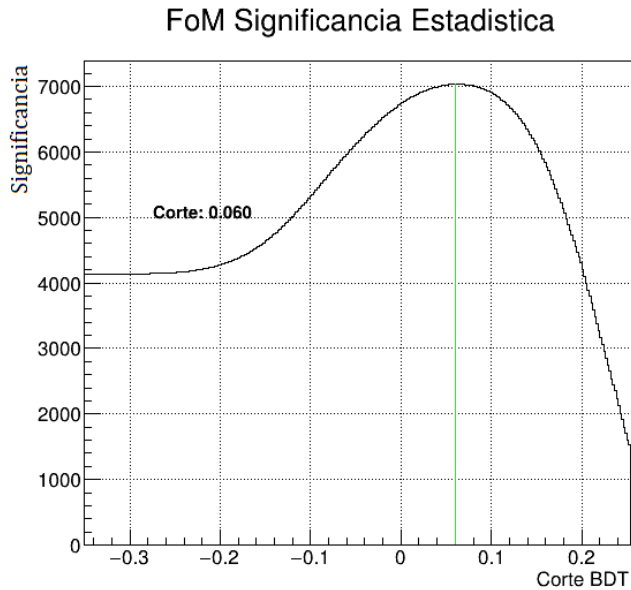


Figura 6.17: FoM de la significancia estadística canal $B^+ \rightarrow K^+ K^- \pi^+$

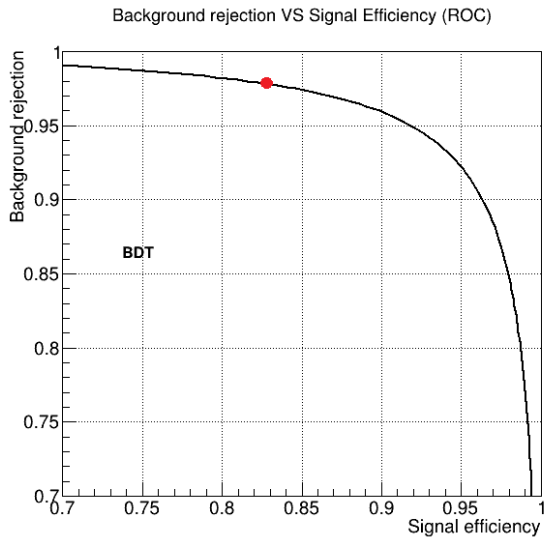


Figura 6.18: Eficiencia de la señal frente al rechazo de background canal $B^+ \rightarrow K^+ K^- \pi^+$

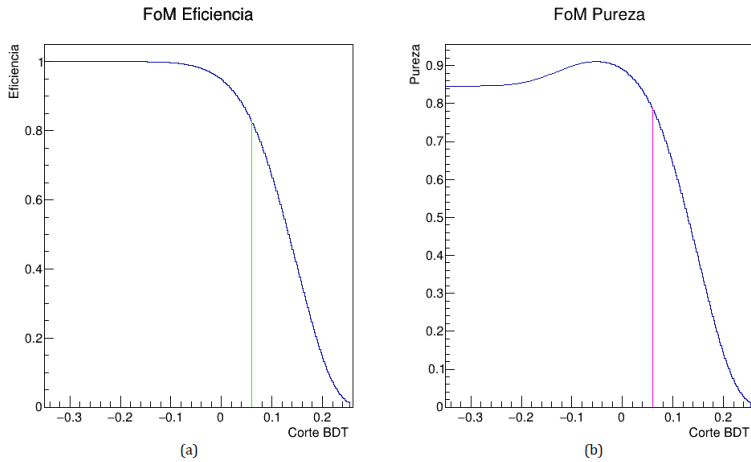


Figura 6.19: En la figura (a): FoM de la eficiencia. (b): FoM pureza

En la Figura 6.18 podemos observar nuevamente la curva ROC solamente de BDT nos da la eficiencia de la señal frente al rechazo de background. Para un 82.86 % de la eficiencia de la señal se logra reducir en un 98 % los eventos de ruido combinatorial, lo cual es otro excelente resultado.

En la Figura 6.19 (a) se muestra la figura de mérito de la eficiencia de la señal. En el corte de $BDT > 0,06$ se logra alcanzar una pureza del 78.76 %, lo cual se visualiza en la Figura 6.19 (b).

6.2.3. Aplicación del corte óptimo de BDT

Ahora se verá la aplicación del corte óptimo de BDT encontrado en la sección 6.2.2 a la distribución de probabilidad de B_m para el canal $B^+ \rightarrow \pi^+\pi^+\pi^-$ y $B^+ \rightarrow K^+K^-\pi^+$ como se puede observar en la Figura 6.20 y 6.21 respectivamente. Se ha aplicado el corte óptimo de $BDT > 0,044$ para el canal $B^+ \rightarrow \pi^+\pi^+\pi^-$ y el corte óptimo de $BDT > 0,06$ para el canal $B^+ \rightarrow K^+K^-\pi^+$. Donde podemos apreciar en ambos canales un pico mayor de eventos de señal, confirmando que se logra una alta reducción del ruido combinatorial que es precisamente uno de los objetivos de este estudio.

6.3. Selección de Identificación de Partículas (PID)

Para tener un criterio de selección completa se realizó una optimización de las variables PID, estas variables son basadas en machine learning. Con el objetivo de obtener una señal utilizable para extraer una medida física. Este es el último paso para reducir el resto del ruido en los cuales predomina el ruido de otros canales hadrónicos (identificación errónea de partículas) y decaimientos parcialmente reconstruidos, esto se describe en la sección 4.5.

6.3.1. Optimización (PID)

Aquí se verá los resultados de lo descrito en las secciones 4.5 y 4.6. Se encontrará el criterio de selección PID para las variables ProbNN en ambos canales hadrónicos. Se comenzará con el análisis de las figuras de mérito (FoM) para las variables ProbNN, en ellas optimizaremos la significancia estadística encontrando el corte PID que disminuya el resto del

CAPÍTULO 6. ANÁLISIS Y RESULTADOS

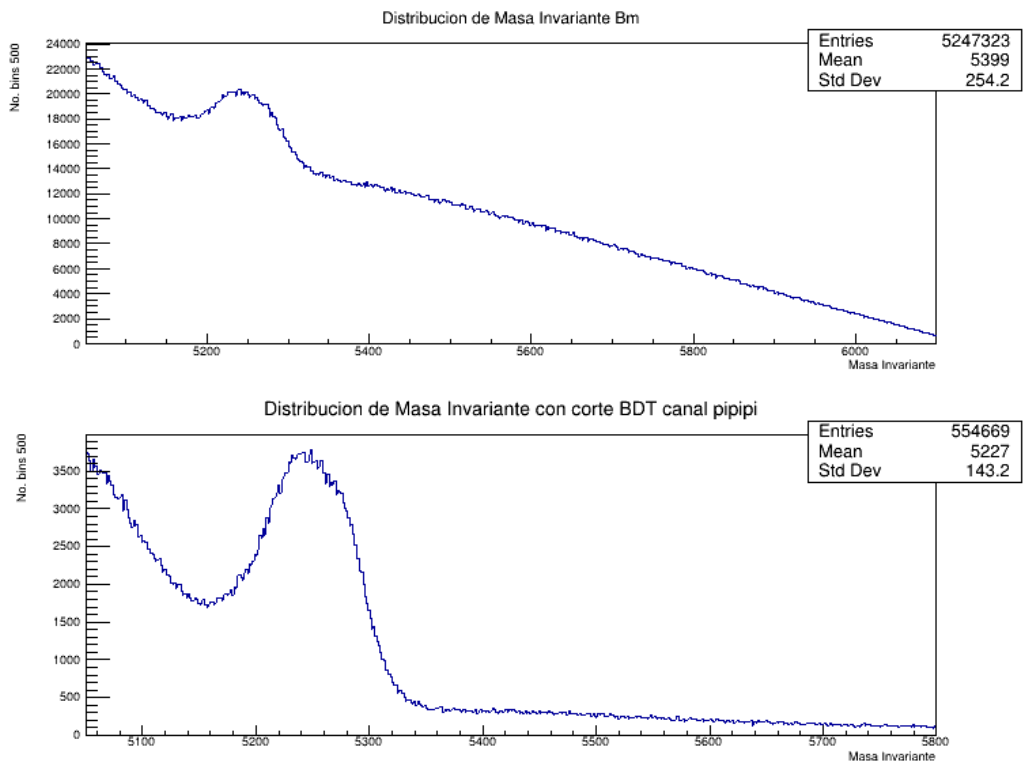


Figura 6.20: Distribución de Probabilidad de B_m (arriba) y Distribución de Probabilidad de B_m con el corte de BDT (abajo), canal $B^+ \rightarrow \pi^+\pi^+\pi^-$

ruido siempre con el compromiso de mantener la mayor eficiencia de la señal.

En las Figuras 6.22, 6.23 y 6.24 se muestran las figuras de mérito de la significancia estadística y la eficiencia de la señal para $d1_ProbNN\pi$, $d2_ProbNN\pi$ y $d3_ProbNN\pi$, respectivamente de $B^+ \rightarrow \pi^+\pi^+\pi^-$.

Podemos observar que para $d1_ProbNN\pi$ la máxima significancia se alcanza en $d1_ProbNN\pi > 0,52$ con una eficiencia de la señal de 93.71 %, para $d2_ProbNN\pi$ la máxima significancia se alcanza en $d2_ProbNN\pi > 0,15$

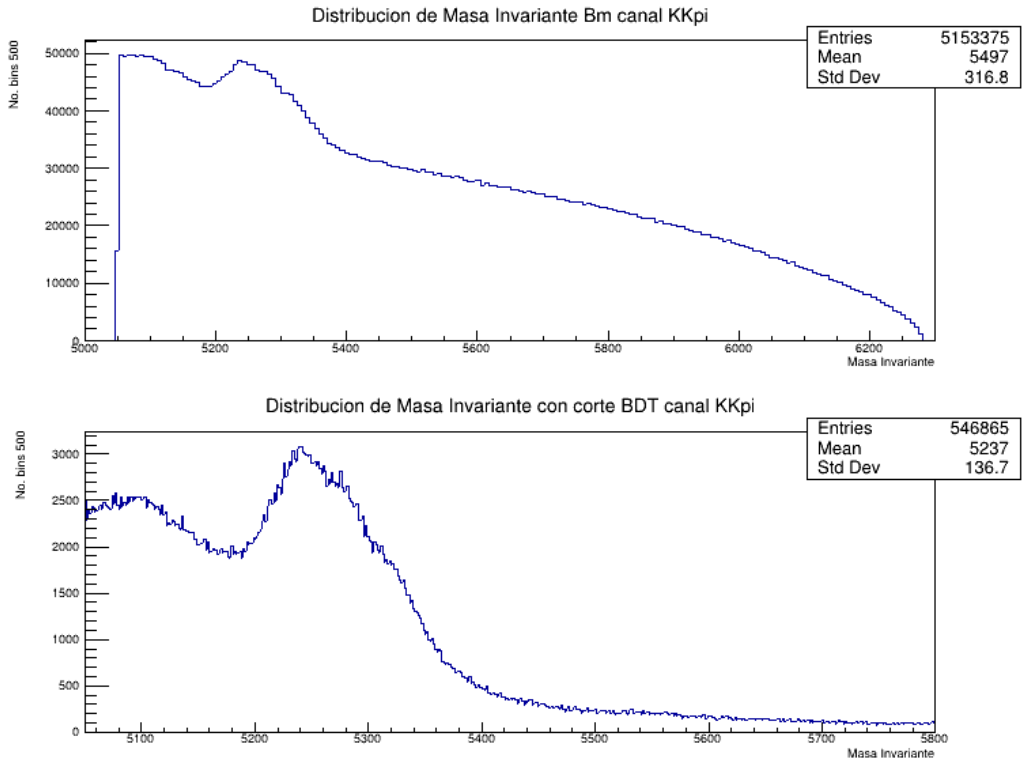


Figura 6.21: Distribución de Probabilidad de B_m (arriba) y Distribución de Probabilidad de B_m con el corte de BDT (abajo), canal $B^+ \rightarrow K^+K^-\pi^+$

con una eficiencia de la señal de 98.37 % y para d3_ProbNNpi la máxima significancia se alcanza en $d3_ProbNNpi > 0,191$ con una eficiencia de la señal de 98.18 %.

Para el decaimiento $B^+ \rightarrow \pi^+\pi^+\pi^-$ se tiene un rechazo de ruido para d1_ProbNNpi, d2_ProbNNpi y d3_ProbNNpi del 30 %, 5 % y del 8 %, respectivamente. Esto se observa mediante las curvas ROC del apéndice B.

Ahora bien, en las Figuras 6.25, 6.26 y 6.27 se muestran las figuras de mérito de la significancia estadística y la eficiencia de la señal para

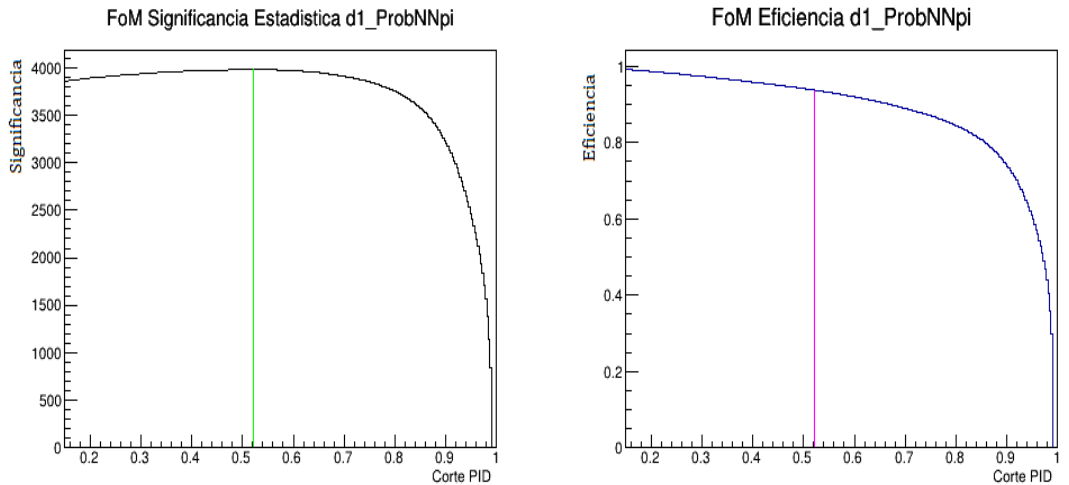


Figura 6.22: Izquierda: FoM significancia estadística, Derecha: FoM Eficiencia, indicando el corte PID para d1_ProbNNpi de $B^+ \rightarrow \pi^+\pi^+\pi^-$

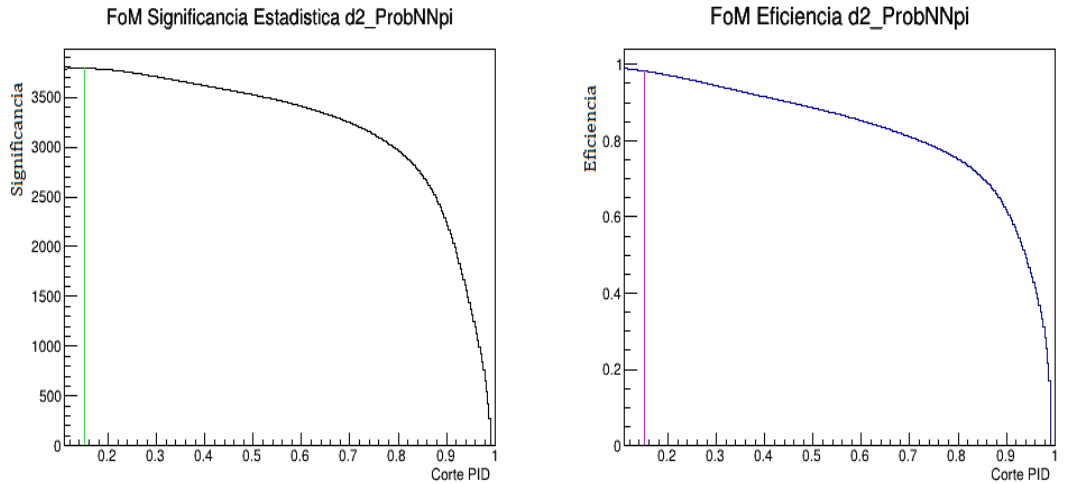


Figura 6.23: Izquierda: FoM significancia estadística, Derecha: FoM Eficiencia, indicando el corte PID para d2_ProbNNpi de $B^+ \rightarrow \pi^+\pi^+\pi^-$

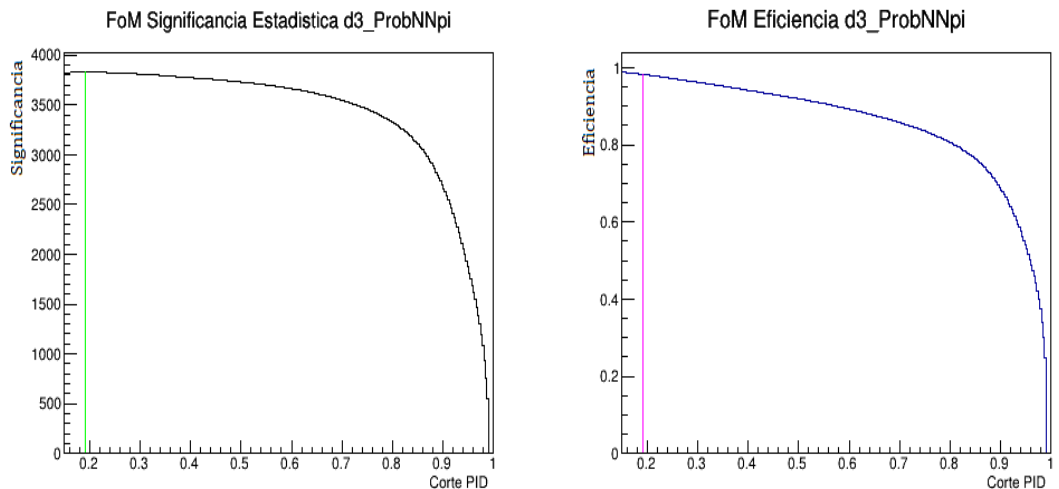


Figura 6.24: Izquierda: FoM significancia estadística, Derecha: FoM Eficiencia, indicando el corte PID para d3_ProbNNpi de $B^+ \rightarrow \pi^+ \pi^+ \pi^-$

d1_ProbNNK, d2_ProbNNpi y d3_ProbNNK, respectivamente de $B^+ \rightarrow K^+ K^- \pi^+$.

Podemos observar que para d1_ProbNNK la máxima significancia se alcanza en $d1_ProbNNK > 0,311$ con una eficiencia de la señal de 90.67 %, para d2_ProbNNpi la máxima significancia se alcanza en $d2_ProbNNpi > 0,266$ con una eficiencia de la señal de 96.52 % y para d3_ProbNNK la máxima significancia se alcanza en $d3_ProbNNK > 0,422$ con una eficiencia de la señal de 86.28 %.

Para el decaimiento $B^+ \rightarrow K^+ K^- \pi^+$ se tiene un rechazo de ruido para d1_ProbNNK, d2_ProbNNpi y d3_ProbNNK del 46 %, 14 % y del 57 %, respectivamente. Esto lo pueden observar mediante las curvas ROC del apéndice B.

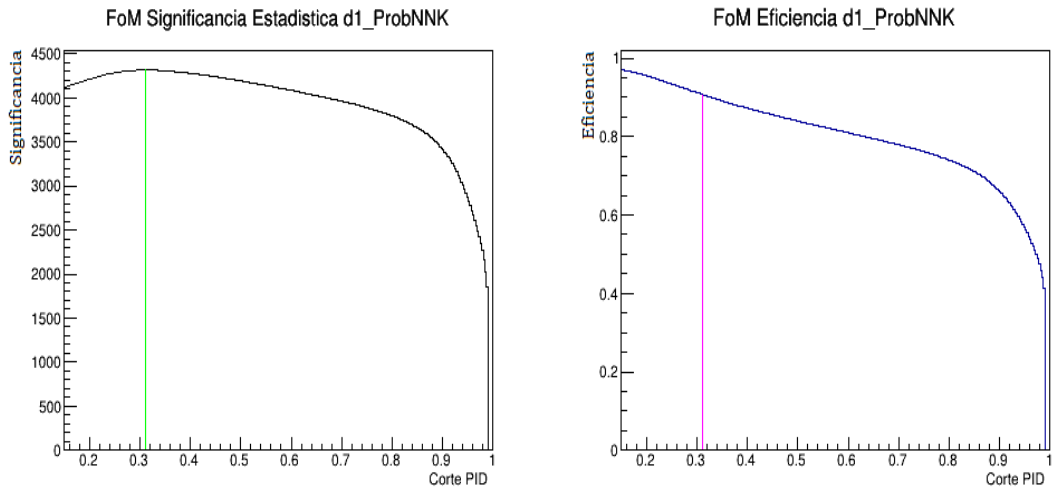


Figura 6.25: Izquierda: FoM significancia estadística, Derecha: FoM Eficiencia, indicando el corte PID para d1_ProbNNK de $B^+ \rightarrow K^+ K^- \pi^+$

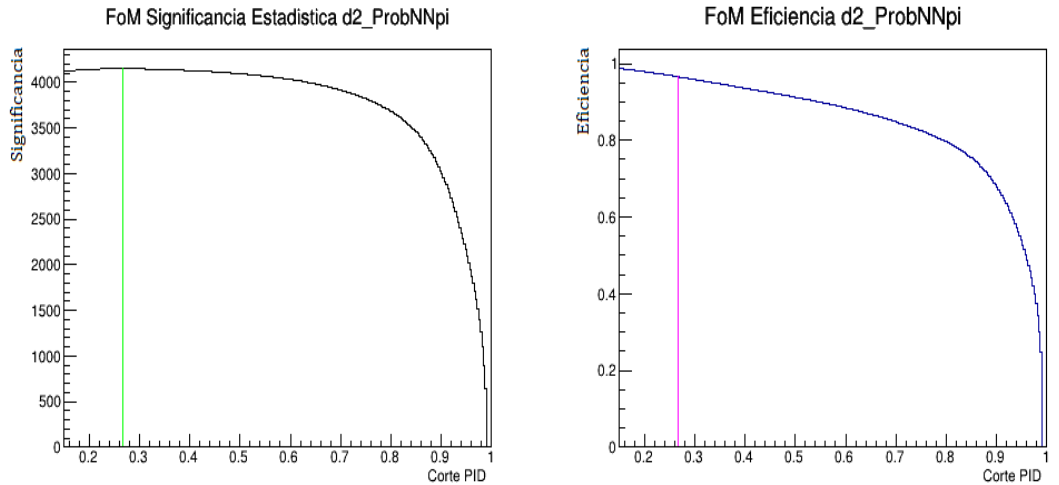


Figura 6.26: Izquierda: FoM significancia estadística, Derecha: FoM Eficiencia, indicando el corte PID para d2_ProbNNpi de $B^+ \rightarrow K^+ K^- \pi^+$

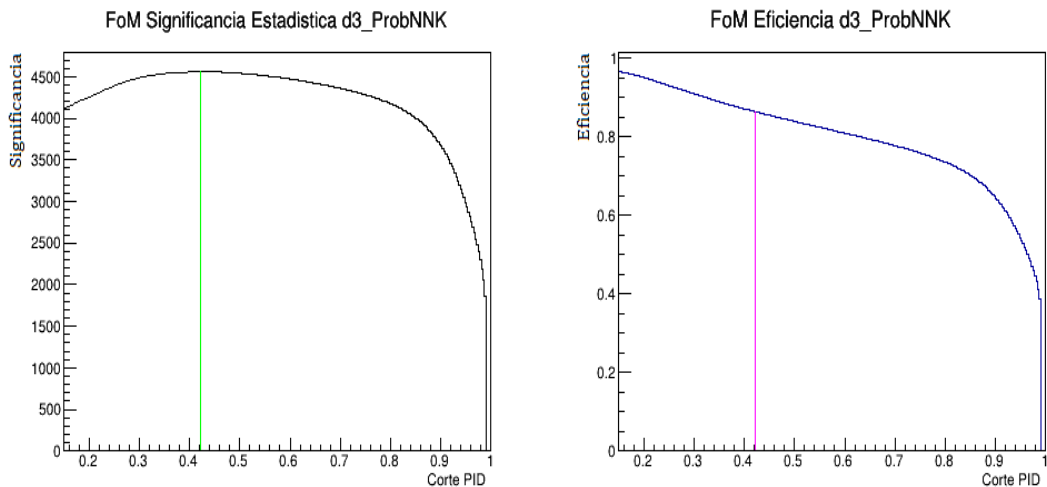


Figura 6.27: Izquierda: FoM significancia estadística, Derecha: FoM Eficiencia, indicando el corte PID para d3_ProbNNK de $B^+ \rightarrow K^+ K^- \pi^+$

6.3.2. Aplicación del criterio de selección PID

Todo el análisis de las figuras de mérito de la sección 6.3.1 se resume en la Tabla 6.3, el cual nos da el criterio de selección PID, que es el último paso del proceso de selección.

Decaimiento	Criterio de Selección PID
$B^+ \rightarrow \pi^+ \pi^+ \pi^-$	$d1_ProbNNpi > 0,521 \ \& \ d2_ProbNNpi > 0,15 \ \& \ d3_ProbNNpi > 0,191$
$B^+ \rightarrow K^+ K^- \pi^+$	$d1_ProbNNK > 0,311 \ \& \ d2_ProbNNpi > 0,266 \ \& \ d3_ProbNNK > 0,422$

Tabla 6.3: Criterio de selección PID

En las Figuras 6.28 y 6.29 se muestra la aplicación del criterio de selección PID a la distribución de masa invariante (B_m) para $B^+ \rightarrow \pi^+ \pi^+ \pi^-$ y $B^+ \rightarrow K^+ K^- \pi^+$, respectivamente. La distribución de masa invariante ya tiene aplicado la selección de análisis multivariante, es decir, ya tiene el corte óptimo de BDT.

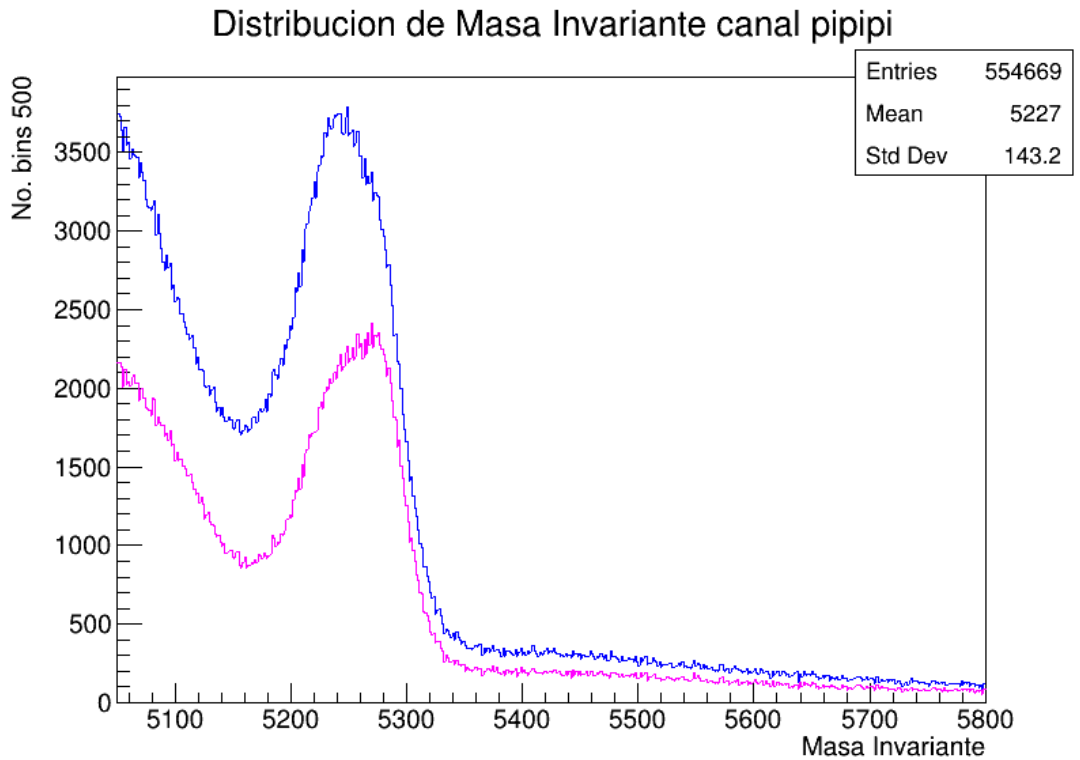


Figura 6.28: Azul: Distribución de probabilidad de B_m con el corte óptimo de BDT. Magenta: Distribución de probabilidad de B_m con el corte óptimo de BDT & PID del decaimiento $B^+ \rightarrow \pi^+\pi^+\pi^-$

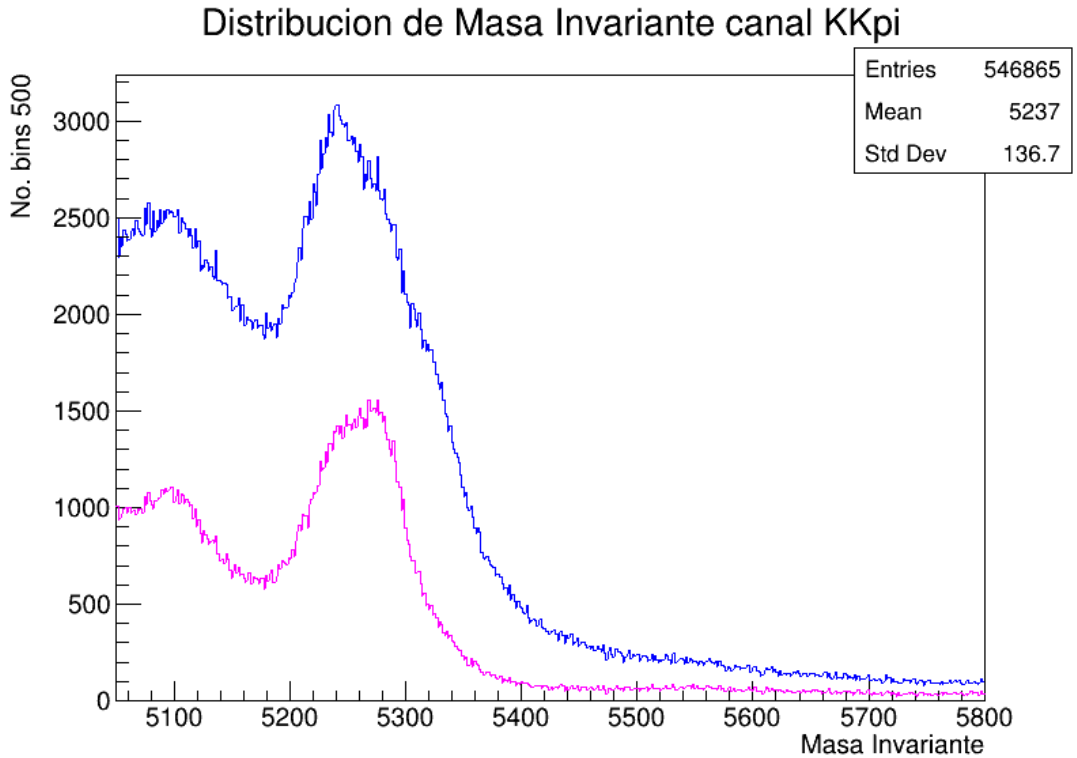


Figura 6.29: Azul: Distribución de probabilidad de B_m con el corte óptimo de BDT. Magenta: Distribución de probabilidad de B_m con el corte óptimo de BDT & PID del decaimiento $B^+ \rightarrow K^+ K^- \pi^+$

Capítulo 7

Conclusiones y Trabajos Futuros

Este trabajo se ha enmarcado en el estudio del decaimiento del mesón B, estas son las conclusiones para los decaimientos $B \rightarrow \pi\pi\pi$ y $B \rightarrow KK\pi$ con datos del 2015 y 2016 del Run 2.

7.1. Conclusiones

- Se utilizó el paquete de TMVA en ROOT para el entrenamiento de BDT y otros clasificadores, encontrando como mejor clasificador al BDT para ambos decaimientos .
- Se observa una alta eficiencia del corte de BDT en la reducción de eventos ruido combinatorial manteniendo la mayor cantidad de eventos de señal. El corte óptimo de BDT para $B \rightarrow \pi^+\pi^+\pi^-$ es $BDT > 0,04$, logrando una eficiencia de la señal de un 88.85 % con un rechazo de ruido del 97 %. El corte óptimo de BDT para $B \rightarrow K^+K^-\pi^+$ es $BDT > 0,06$, logrando una eficiencia de la señal de un 82.85 % con un rechazo de ruido del 98.1 %. Para ambos decaimientos en el corte óptimo se alcanza la máxima significancia estadística y la máxima pureza. Lo cual es un excelente resultado, tanto en la mis-identificación de ruido combinatorial como en la optimización de las figuras de mérito.

- Con la selección de identificación de partículas (PID), se logra disminuir el resto del ruido. Cuando se aplica el criterio de selección PID para $B \rightarrow \pi^+\pi^+\pi^-$ se aprecia que hay una reducción de ruido de mis-identificación y comparado con el corte de BDT, la distribución de los datos tiene un comportamiento similar. De igual forma, para $B \rightarrow K^+K^-\pi^+$ se aprecia que hay una reducción de ruido de mis-identificación y comparado con el corte de BDT la diferencia se ve mayor, sin embargo, la distribución propia de los datos mejoró.
- Estos resultados constituyen un escenario se pueden investigar muchos más.
- La importancia de trabajar con estos canales en específico los decaimientos de tres cuerpos de mesones B^+ es que son laboratorios para medidas de violación CP, por lo tanto, son canales de alta relevancia para la comunidad experimental y teórica de Física de Partículas. Estos decaimientos de tres cuerpos son procesos poco frecuentes y se necesitan de grandes aceleradores de partículas para poder estudiarlos. Desde el punto de vista físico lo que se busca con el estudio de estos canales es entender la diferencia que hay entre la materia y la antimateria en el universo. Nuestro universo es dominado por la materia, al inicio se debe haber formado un exceso de materia para que la misma sobreviviera a la aniquilación. Para producir tal exceso, debe existir una diferencia entre la materia y la antimateria y es aquí donde entra la asimetría CP que nos ayuda a entender esa diferencia.

7.2. Trabajos Futuros

- Aplicación de la técnica como herramienta para la exploración de datos de naturaleza diferente a los presentados en esta disertación.
- Estudiar los tipos de ruidos en decaimientos diferentes a los estudiados en esta disertación. Usualmente cada decaimiento posee su propia y compleja contaminación.
- Utilizar otras técnicas de Machine Learning como redes neuronales para la reducción del ruido combinatorial y comparar los rendimientos con los obtenidos en esta disertación.

CAPÍTULO 7. CONCLUSIONES Y TRABAJOS FUTUROS

- Realizar el estudio con la muestra de datos completa del Run 2 e identificar las diferencias con esta investigación que involucra los años 2015 y 2016.
- Identificar otras variables que sean buenas discriminantes entre señal y ruido, volver a realizar el análisis y hacer comparaciones con los resultados ya obtenidos.
- Exploración de los diferentes tipos de ruidos con la aplicación de discriminantes alternativos para realizar estudios de errores sistemáticos.

Apéndice A

Análisis Multivariante

A.1. Separación de las variables

Rango	Variable	Separación
1	pointing_pt	0.48
2	ipownpvsum	0.34
3	B_AMAXDOCA	0.27
4	B_ENDVERTEX_CHI2	0.17
5	ptmed	0.16
6	B_050_cc_CONEPTASYM_B	0.15
7	B_P	0.12
8	ptmax	0.11
9	d3_P	0.02
10	B_IPCHI2_OWNPV	0.009

Tabla A.1: Variables discriminantes utilizadas en el entrenamiento de MVA, ordenadas de-decientemente según su poder de separación para el decaimiento $B \rightarrow \pi^+ \pi^+ \pi^-$

Rango	Variable	Separación
1	pointing_pt	0.47
2	B_AMAXDOCA	0.29
3	B_050_cc_CONEPTASYM.B	0.19
4	B.P	0.187
5	ipownpvsum	0.18
6	B_ENDVERTEX_CHI2	0.16
7	ptmed	0.15
8	d3_P	0.09
9	ptmax	0.038
10	B_IPCHI2_OWNPV	0.036

Tabla A.2: Variables discriminantes utilizadas en el entrenamiento de MVA, ordenadas de-crecientemente según su poder de separación para el decaimiento $B \rightarrow K^+ K^- \pi^+$

A continuación definiremos cada una de las variables discriminantes utilizadas en el entrenamiento de MVA, según el orden descrito en el decaimiento $B \rightarrow \pi^+ \pi^+ \pi^-$:

- **pointing_pt** ($pointing_{PT}$):

$$pointing_{PT} = \frac{P_B \sin(\theta)}{P_B \sin(\theta) + \sum_i P_T^i} \quad (\text{A.1.1})$$

donde, θ es el ángulo entre el vector de momento B y la dirección de vuelo. P_B es el momento de B y $\sum_i P_T^i$ es la suma del momento transversal de todas las trayectorias del decaimiento, es decir, es la suma escalar de las hijas de P_T .

- **ipownpvsum** (IP sum): Suma del parámetro de impacto de las trayectorias.
- **B_AMAXDOCA** (MAXDOCA): Valor máximo de la variable Distancia de Aproximación Cercana (DOCA). Donde la variable DOCA se define como la distancia de máxima de aproximación entre dos vías.

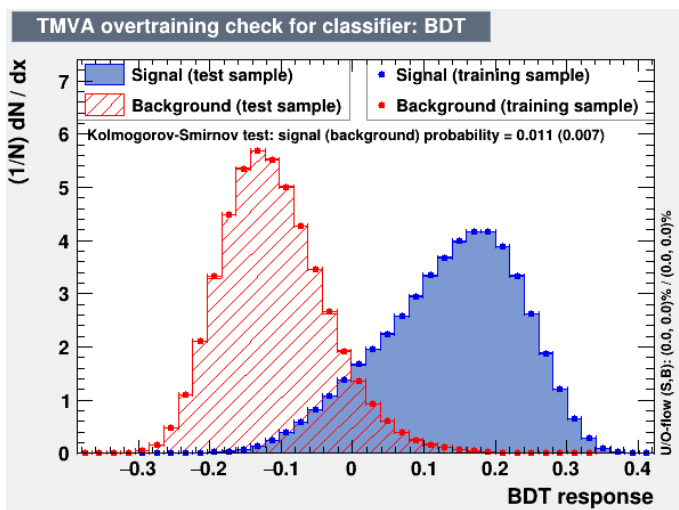
- **B_ENDVERTEX_CHI2** (B vertex χ^2): χ^2 asociado a la reconstrucción del vértice B.
- **ptmed** (P_T^{med}): Es el segundo momento transversal más grande entre las tres trayectorias hijas.
- **B_050_cc_CONEPTASYM_B** ():
- **B_P** (BP): Momento del candidato B
- **ptmax** (P_T^{max}): Máximo momento transversal entre las tres trayectorias hijas
- **d3_P** (d_3 P): Momentum de la tercera hija
- **B_IPCHI2_OWNPV** (B IP χ^2): χ^2 asociado con el parámetro de impacto del candidato B.

A.2. Overtraining

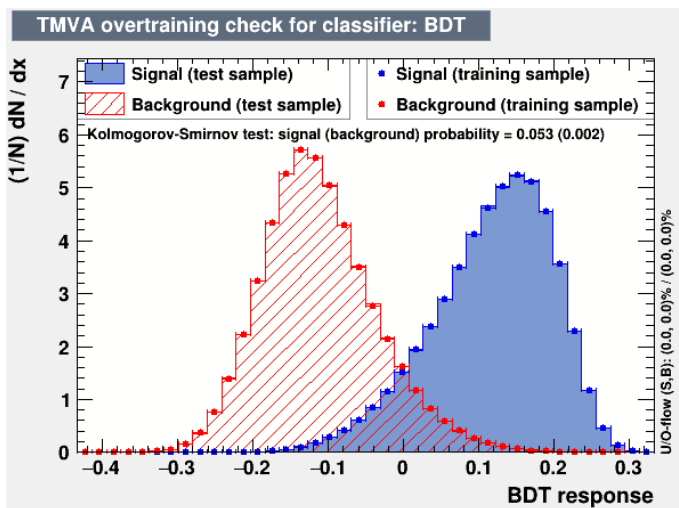
Para verificar si existe sobreentrenamiento (overtraining) en el algoritmo entrenado de BDT, esto se puede cuantificar en un test estadístico (test Kolmogorov-Smirnov, K-S) en nuestras muestras de prueba y entrenamiento. También podemos observar el sobre-entrenamiento en la diferencias de formas que presentan las muestras de entrenamiento y prueba para los eventos de señal y ruido por separado. Aquí podemos decir que la aparición de una gran cantidad de puntos alejados entre sí en los resultados de las muestras es una indicación que estamos es presencia de overtraining.

Las Figuras A.1 muestran los resultados del overtraining de BDT para ambos canales, en los cuales no se observa overtraining.

El valor p nos indica la probabilidad de rechazar o aceptar la hipótesis nula del test K-S. Este test se utiliza en la salida de BDT de una muestra de entrenamiento y una muestra de prueba, con la suposición de que deberían producir el mismo resultado. Por lo tanto, nos interesa que los valores de p sean los mas elevados posibles. Sin embargo, hay que tomar en cuenta que la presencia de las incertidumbres experimentales, que el test estadístico (K-S) no toma en cuenta, nos lleva a que sea muy complicado obtener valores altos en este test de hipótesis.



(a) Decaimiento $B \rightarrow \pi^+ \pi^+ \pi^-$



(b) Decaimiento $B \rightarrow K^+ K^- \pi^+$

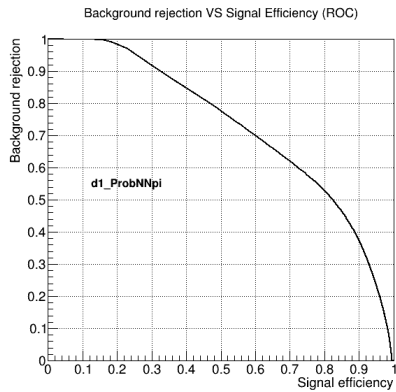
Figura A.1: Verificación del overtraining

Apéndice B

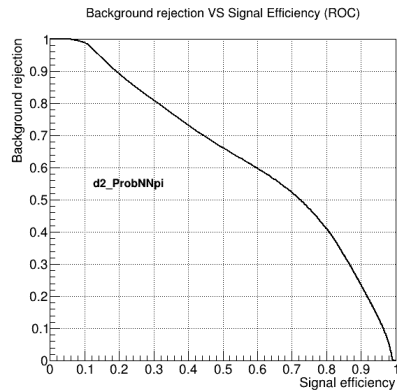
Curvas ROC - Identificación de Partículas

En las Figuras B.1 y B.2 se muestran las curvas ROC para las variables de identificación de partículas ProbNN de los canales $B^+ \rightarrow \pi^+\pi^+\pi^-$ y $B^+ \rightarrow K^+K^-\pi^+$, respectivamente.

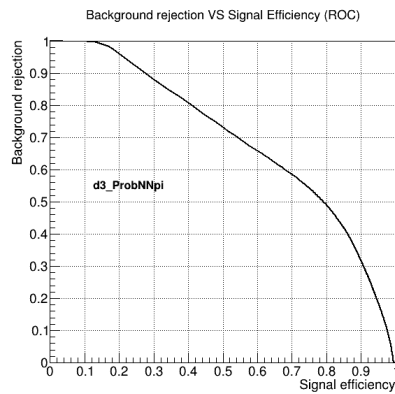
APÉNDICE B. CURVAS ROC - IDENTIFICACIÓN DE PARTÍCULAS



(a) d1_ProbNNpi



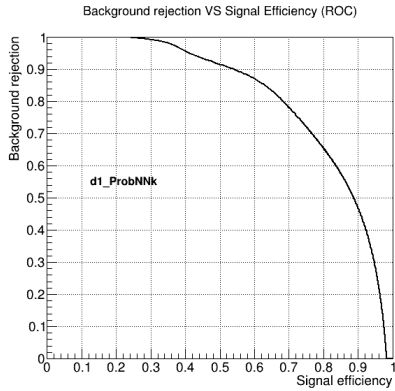
(b) d2_ProbNNpi



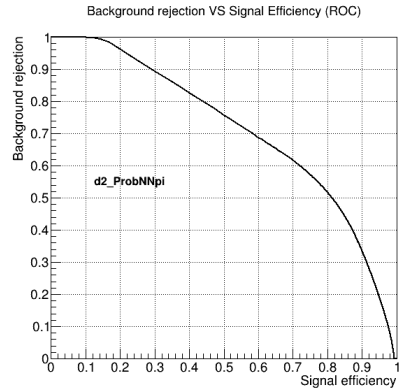
(c) d3_ProbNNpi

Figura B.1: Curvas ROC de variables ProbNN para el canal $B^+ \rightarrow \pi^+ \pi^+ \pi^-$

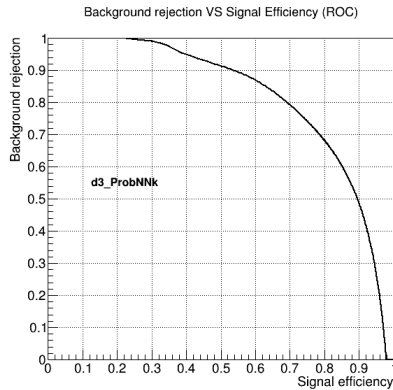
APÉNDICE B. CURVAS ROC - IDENTIFICACIÓN DE PARTÍCULAS



(a) d1_ProbNNk



(b) d2_ProbNNpi



(c) d3_ProbNNk

Figura B.2: Curvas ROC de variables ProbNN para el canal $B^+ \rightarrow K^+ K^- \pi^+$

Bibliografía

- [1] <https://twiki.cern.ch/twiki/bin/view/lhcb/pidcalibpackage>, 4.6.
- [2] Jr. et al. (LHCb collaboration) A. A. Alves. The lhcb detector at the lhcb. *Instrum. 3, S08005*, 2018, 3.1, 3.3, 3.4.3.
- [3] J. Stelzer J. Therhaag E. von Toerne H. Voss A. Hoecker, P. Speckmayer. TMVA 4: Toolkit for Multivariate Data Analysis with ROOT Users Guide. 2013, 5.3, 5.4.3, 5.5, 5.4.1.
- [4] J. Zhu Y. Liu I. Stancu G. McGregor B. P. Roe, H.J. Yang. Boosted Decision Trees as an Alternative to Artificial Neural Networks for Particle Identification. 2004, 5.4.1.
- [5] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Information science and statistics. Springer, 1st ed. 2006. corr. 2nd printing edition, 2006, 2.3.1.
- [6] L. et al BREIMAN. Classification and regression trees. *Belmont, California, USA: Wadsworth international group*, 1984, 5.4.1.
- [7] CERN. <https://home.cern/science/accelerators/large-hadron-collide>, 3.1.
- [8] David Ciupke. Study of BDT Training Configurations with an Application to the $Z/H \rightarrow \tau\tau \rightarrow ee$ Analysis. *University of Göttingen, Germany*, 2012, 5.4.1.
- [9] The LHCb collaboration. LHCb magnet. Technical Report CERN-LHCC-2000-007,CERN, Geneva. *LHCb-TDR-001*, 2000, 3.4.2.

BIBLIOGRAFÍA

- [10] The LHCb collaboration. LHCb RICH. Technical Report CERN-LHCC-2000-037, CERN, Geneva. *LHCb-TDR-003*, 2000, 3.4.4.
- [11] The LHCb collaboration. LHCb calorimeters. Technical Report CERN-LHCC- 2000-036, CERN, Geneva. *LHCb-TDR-002*, 2000, 3.4.5.
- [12] The LHCb collaboration. LHCb VELO (VERTex LOcator). Technical Report CERN- LHCC-2001-011, CERN, Geneva. *LHCb-TDR-005*, 2001, 3.4.1.
- [13] The LHCb collaboration. LHCb outer tracker. Technical Report CERN-LHCC- 2001-024, CERN, Geneva. *LHCb-TDR-006*, 2001, 3.4.3.
- [14] The LHCb collaboration. LHCb muon system. Technical Report CERN-LHCC- 2001-010, CERN, Geneva. *LHCb-TDR-004*, 2001, 3.4.6.
- [15] The LHCb collaboration. LHCb inner tracker. Technical Report CERN-LHCC- 2002-029, CERN, Geneva. *LHCb-TDR-008*, 2002, 3.4.3.
- [16] The LHCb collaboration. LHCb Trigger and Online. Technical Report CERN-LHCC-2014-016, CERN, Geneva. *LHCb-TDR-016*, 2014, 4.3.1, 3.5.
- [17] The LHCb collaboration. Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC. *CERN-LHCb-DP-2019-001*, 2019, 4.3.1, 3.5.
- [18] Glen Cowan. *Statistics for Searches at the LHC*. <https://doi.org/10.48550/arXiv.1307.2487>, 2013, 2.3.1, 2.3.2.
- [19] CPAN. <https://www.i-cpan.es/es/content/¿que-es-la-física-de-partículas>, 2010, 2.2.
- [20] R. Aaij et al. Precision luminosity measurements at LHCb. *Journal of Instrumentation*, *9:P12005*, 2014, 3.2.

BIBLIOGRAFÍA

- [21] R. Aaij et al. Performance of the LHCb Vertex Locator. *Journal of Instrumentation*, 9:P09007, 2014, 3.4.1.
- [22] R. Aaij et al. LHCb detector performance. *International Journal of Modern Physics*, A30:1530022, 2015, 3.1, 3.3.
- [23] G. Barrand et al. GAUDI. The Software Architecture and Framework for building LHCb Data Processing Applications. 2000, 3.6.
- [24] Philippe Charpentier Markus Frank Patrick Koppenburg Pere Mato Florence Ranjard Stefan Roiser Ivan Belyaev Gloria Corti, Marco Cattaneo and Guy Barrand. Software for the LHCb Experiment. *IEEE TRANSACTIONS ON NUCLEAR SCIENCE*, 53, 2006, 3.6.
- [25] D. Griffiths. *Introduction to Elementary Particles*. Wiley-VCH, 2 edition edition, 2008, 2.2.
- [26] Particle Data Group. <https://pdg.lbl.gov/2021/tables/contents-tables.html>, 2.2.
- [27] Mario Señas Gómez. Simulaciones realistas de colisiones protón-protón en el lhc usando una red neuronal convolucional extractora de correlaciones locales. 1, 8 2018, 5.1.
- [28] Alejandro De Yta Hernández. Estudio de técnicas multivariadas para la búsqueda de decaimientos $B^0 \rightarrow D^- \pi^- \mu^+ \mu^+$ en el experimento Belle II, 2016. 5.4.1.
- [29] G Kane. The dawn of physics beyond the standard model. *Scientific American*, 2003, 2.2.
- [30] C.R. Jones S. Malde D. Mülle5 S. Ogilv6 J.M. Otalora Goicochea A. Pearce I. Polyakov W. Qian B. Sciascia R. Vazquez Gomez Y. Zhang L. Anderlini, A. Contu. The PIDCalib package. *LHCb-PUB-2016-021*, 2016, 4.6.
- [31] LHCb. <https://lhcb-outreach.web.cern.ch/2022/03/17/magic-beauty-charmless-decays/>, 2.2.
- [32] Luca Lista. *Statistical Methods for Data Analysis in Particle Physics*. Lecture Notes in Physics 909. Springer, 2016 edition, 2015, 2.3.1.

BIBLIOGRAFÍA

- [33] Julian Stenzel Martins. Contribution to the Analysis of the $B_c^+ \rightarrow K^+ K^- \pi^+$ Decay in the LHCb Experiment, 2016. 5.4.1.
- [34] Rick Mijsbergh. Classifying photons with machine learning in ALICE, 2016. 5.4.1.
- [35] LHCb Collaboration R. Aaij et al. Measurements of CP violation in the three-body phase space of charmless B^\pm decays. 2014, 1, 2.1.
- [36] CERN. ROOT. a Data Analysis Framework, <https://root.cern.ch/>, 5.3.
- [37] J. H. Friedman T. Hastie, R. Tibshirani. *The Elements of Statistical Learning*. Springer, corrected edition, 2003, 5.4.1.
- [38] Melissa María Cruz Torres. Study of CP violation and amplitude analysis of the decay $B^+ \rightarrow \pi^+ K^- K^+$ in the LHCb experiment, 2017. 4.3.1, 5.4.1.
- [39] Ramón Cid Manzano Xabier Cid Vidal. Luminosidad: la luz al final del túnel (lhc). *Lat. Am. J. Phys. Educ*, Vol. 3, 2009, 3.2.