

MAYA: ***MA**chine learning **pY**thon **A**pplication for HEP.*

Presentador: Marco Antonio Arroyo Ureña

CIIEC, FCFM-BUAP

September 29, 2025

Desarrolladores¹

Manuel Pérez-Ortega, Diego Carreño-Díaz,



Edwin Herrera-Chacón, Marco A. Arroyo-Ureña.



¹Agradecimientos al Dr. Jhovanny Mejía por sus valiosos comentarios. ▶

Contenido

- ▶ ¿Qué es MAYA?



- ▶ Motivación
- ▶ Introducción
- ▶ Ejemplo: Búsqueda del Flavon en el canal difotónico ($pp \rightarrow H_F \rightarrow \gamma\gamma$).
- ▶ Conclusiones y perspectivas

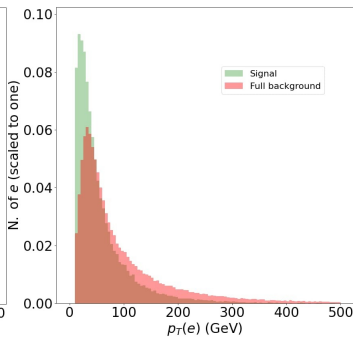
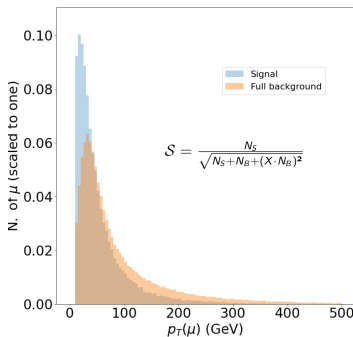
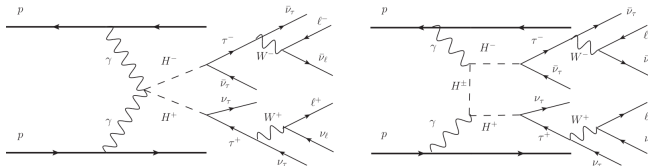
māyā



- ▶ MAYA es una herramienta diseñada en `python` para el análisis de eventos de colisionadores a través de una interfaz gráfica. Integra el método de árboles de decisión “potenciados” mediante la biblioteca XGBoost para la clasificación señal-ruido de fondo, optimiza hiperparámetros usando Optuna, y visualiza resultados con histogramas. Como principal objetivo, MAYA automatiza la predicción de significancias estadísticas.
- ▶ La significancia estadística $\mathcal{S} = \frac{N_S}{\sqrt{N_S + N_B + (X \cdot N_B)^2}}$ mide qué tan improbable es que una señal observada sea producto de una fluctuación del ruido de fondo (background), en lugar de ser una señal real de nueva física: $3\sigma \rightarrow$ **EVIDENCIA**, $5\sigma \rightarrow$ **DESCUBRIMIENTO**.

Motivación

- Colisión ultraperiférica² $pp \rightarrow p(\gamma\gamma \rightarrow H^- H^+)p$



Motivación

► Colisiones protón-protón³ $pp \rightarrow th, (t \rightarrow \ell \nu_\ell, h \rightarrow \gamma\gamma)$.

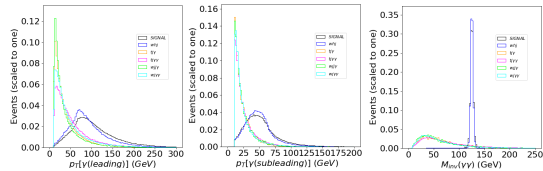
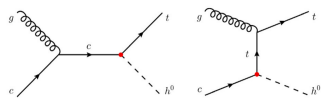


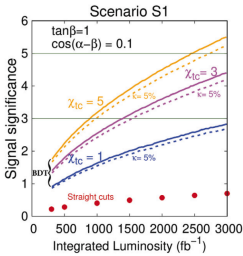
Table 3
Cross-section (σ) of the signal for the scenarios S1, S2.

Scenario	σ [fb] ($\chi_{tc} = 1$)	σ [fb] ($\chi_{tc} = 5$)
S1	0.01	0.025
S2	0.004	0.014

$$S = \frac{N_S}{\sqrt{N_S + N_B + (X \cdot N_B)^2}}$$

Table 4
Cross-section of the dominant SM background processes.

SM backgrounds	Cross-section [fb]
$pp \rightarrow Whj$	0.19
$pp \rightarrow W^j\gamma\gamma$	68.27
$pp \rightarrow t\bar{t}\gamma\gamma$	0.019
$pp \rightarrow W^jj\gamma$	16590
$pp \rightarrow t\bar{t}j\gamma$	1.84



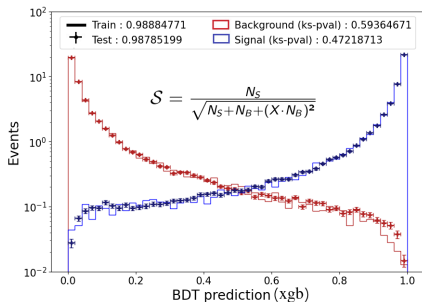
} māyā



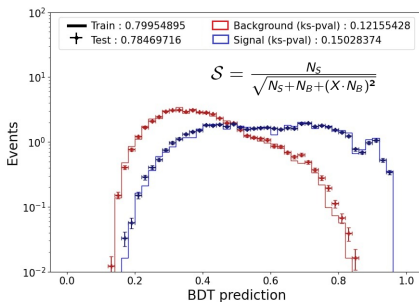
Motivación

Figure: Clasificador

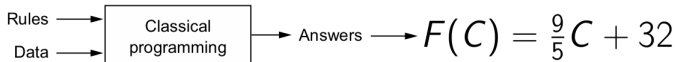
$pp \rightarrow th, (t \rightarrow \ell \nu_\ell b, h \rightarrow \gamma\gamma)$



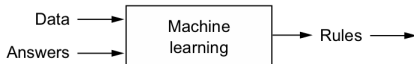
$pp \rightarrow p(\gamma\gamma \rightarrow H^- H^+)p$



Introducción

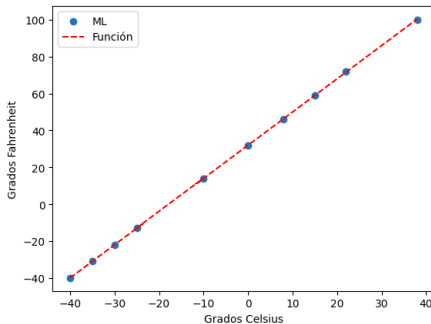


```
celsius = np.array([-40, -35, -30, -10, -25, 0, 8, 15, 22, 38], dtype=float)
fahrenheit = np.array([-40, -31, -22, 14, -13, 32, 46, 59, 72, 100], dtype=float)
```



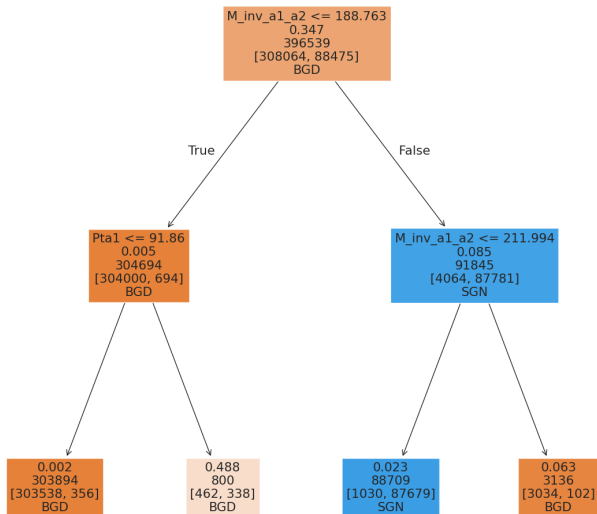
```
print("Variables internas del modelo")
print(capa.get_weights())

Variables internas del modelo
[array([[1.7981262]]), dtype=float32], array([[31.949242]], dtype=float32)]
```



Introducción

► Árbol de decisión



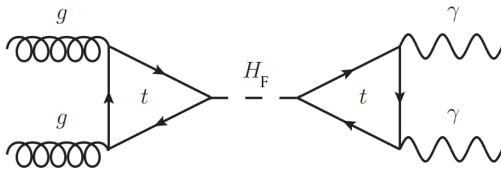
Introducción

1. Lagrangiano $\rightarrow \left\{ \begin{array}{l} \text{LanHEP (Lenguaje C)} \\ \text{FeynRules (Mathematica)} \\ \text{Sarah (Mathematica)} \end{array} \right.$,
2. Salidas: $\left\{ \begin{array}{l} \text{CalcHEP, MicrOmegas} \rightarrow .mdl \\ \text{MadGraph, MadDM} \rightarrow .py \\ \text{FeynArts} \rightarrow .mod \end{array} \right. \rightarrow \sigma, \mathcal{BR}, \Omega, \textbf{EVENTOS}$
3. **EVENTOS** \rightarrow lhe, hepmc, root, lhco \rightarrow **bases de datos.**

Ejemplo: $H_F \rightarrow \gamma\gamma$ ⁴

Búsqueda del Flavon en el canal difotónico⁵

Figure: Diagrama de Feynman del proceso $pp \rightarrow H_F \rightarrow \gamma\gamma$.

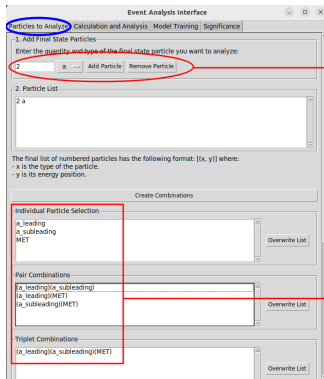


⁴Phys. Rev. D 111, 115037

⁵Ver charla de Diego Carreño: miércoles 9:50 hrs.

Ejemplo: $H_F \rightarrow \gamma\gamma$

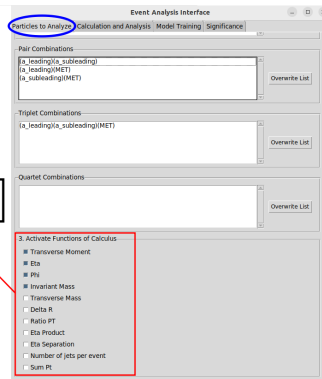
Figure: Interfaz de MAYA



Estado final de la señal.

Observables para entrenar y graficar.

Combinación de características



Ejemplo: $H_F \rightarrow \gamma\gamma$

Figure: Interfaz de MAYA

The screenshot shows the MAYA Event Analysis Interface with several sections and annotations:

- Top Bar:** Contains tabs for "Particles to Analyze", "Calculation and Analysis" (circled in blue), "Model Training", and "Significance".
- 1. Upload Signal Files:** A text input field with the annotation "Buscar, eliminar, cargar datos SEÑAL" and a button group with "Browse File", "Remove File", and "Load Signal".
- Upload Background Files:** A text input field with the annotation "Buscar, eliminar, cargar datos BACKGROUND" and a button group with "Browse File", "Remove File", and "Load Background".
- 2. Event Processing:** Includes a "Filter Events" section and a "Start Calculation" button.
- 3. Select Histogram Column:** A section with the instruction "Select the column you want to display as a histogram once the observables have been calculated." and a dropdown menu. An annotation "Filtración y creación de base de datos para entrenamiento/prueba" points to this section.
- 4. Plot Customization:** A large section with fields for "Custom title", "Legend title", "X-axis label", "Y-axis label", "X Range (min,max)", "Number of bins", "Logarithmic Scale", and "Fit type". A red box encloses this entire section, with an annotation "Gráficación" pointing to it. A "Generate Plot" button is located at the bottom right of this section.

Ejemplo: $H_F \rightarrow \gamma\gamma$

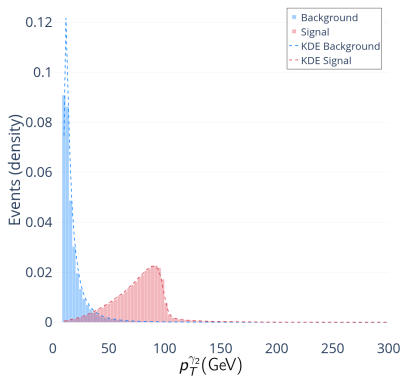
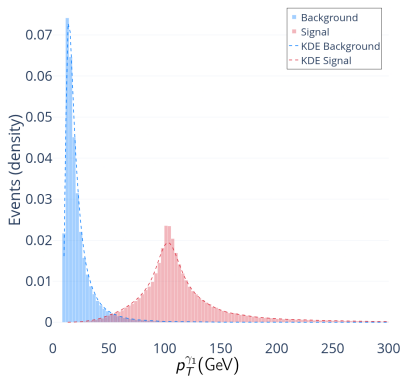
Figure: root - lhco

[illegible]

	typ	eta	phi	pt	jnas	ntrk	btag	had/em	dun1	dun2
0	0	0	0	0	0	0	0	0	0	0
1	0	-0.496	1.673	125.56	0.00	0.0	0.0	0.00	0.0	0.0
2	4	-3.152	-2.177	67.88	-0.00	0.0	0.0	0.00	0.0	0.0
3	4	-1.553	-0.476	38.68	7.50	9.0	0.0	0.00	0.0	0.0
4	6	0.000	-1.111	34.77	0.00	0.0	0.0	0.00	0.0	0.0
0	0	1	0	0	0	0	0	0	0	0
1	4	2.150	0.876	158.01	1.55	1.0	0.0	0.00	0.0	0.0
2	4	-0.550	-2.555	65.93	10.86	3.0	0.0	0.04	0.0	0.0
3	4	-1.981	-1.464	63.86	16.05	5.0	0.0	0.18	0.0	0.0
4	4	3.070	-2.397	22.86	3.97	0.0	0.0	3.83	0.0	0.0
5	6	0.000	0.793	19.67	0.00	0.0	0.0	0.00	0.0	0.0
0	0	2	0	0	0	0	0	0	0	0
1	0	-2.170	-1.792	276.22	0.00	0.0	0.0	0.00	0.0	0.0
2	1	-2.053	0.050	182.41	0.00	1.0	0.0	0.00	0.0	0.0
3	1	-2.026	0.163	38.20	0.00	-1.0	0.0	0.00	0.0	0.0
4	4	0.080	1.657	222.64	14.28	11.0	0.0	0.10	0.0	0.0
5	4	-2.569	-3.124	119.92	19.52	1.0	0.0	2.48	0.0	0.0
6	6	0.000	2.312	31.06	0.00	0.0	0.0	0.00	0.0	0.0
0	0	3	0	0	0	0	0	0	0	0
1	0	-0.692	-1.615	10.00	0.00	0.0	0.0	0.00	0.0	0.0
2	4	-1.834	1.370	1.00	0.00	0.0	0.0	0.02	0.0	0.0
3	6	0.000	0.194	1.00	0.00	0.0	0.0	0.00	0.0	0.0
0	0	4	0	0	0	0	0	0	0	0
1	0	0.786	-1.300	199.54	0.00	0.0	0.0	0.00	0.0	0.0
2	0	1.437	1.635	180.03	0.00	0.0	0.0	0.00	0.0	0.0
3	4	-1.042	-2.835	80.99	15.55	14.0	0.0	0.49	0.0	0.0
4	4	-1.930	0.422	59.63	10.18	9.0	0.0	0.92	0.0	0.0
5	4	-2.805	0.513	33.87	7.29	0.0	0.0	3.63	0.0	0.0
6	4	-0.156	-2.802	26.82	8.99	7.0	0.0	0.00	0.0	0.0
7	6	0.000	1.705	22.69	0.00	0.0	0.0	0.00	0.0	0.0
0	0	5	0	0	0	0	0	0	0	0
1	0	-2.245	-2.720	178.85	0.00	0.0	0.0	0.00	0.0	0.0
2	0	-1.148	0.433	178.04	0.00	0.0	0.0	0.00	0.0	0.0
3	6	0.000	-1.791	9.52	0.00	0.0	0.0	0.00	0.0	0.0
0	0	6	0	0	0	0	0	0	0	0
1	0	-2.286	-1.132	167.37	0.00	0.0	0.0	0.00	0.0	0.0
2	0	-1.023	2.228	165.92	0.00	0.0	0.0	0.00	0.0	0.0
3	4	-3.304	0.666	31.93	7.85	0.0	0.0	10.46	0.0	0.0
4	4	0.174	2.300	23.14	7.33	8.0	0.0	0.00	0.0	0.0
5	6	0.000	-0.654	20.49	0.00	0.0	0.0	0.00	0.0	0.0
0	0	7	0	0	0	0	0	0	0	0
1	0	0.151	0.729	164.42	0.00	0.0	0.0	0.00	0.0	0.0
2	4	1.539	-2.860	165.98	0.00	0.0	0.0	0.00	0.0	0.0
3	4	0.282	-1.251	61.85	9.88	7.0	0.0	1.64	0.0	0.0
4	6	0.000	-1.318	6.79	0.00	0.0	0.0	0.00	0.0	0.0
0	0	8	0	0	0	0	0	0	0	0
1	0	0.234	0.853	221.96	0.00	0.0	0.0	0.00	0.0	0.0
2	0	0.319	-2.020	184.40	0.00	0.0	0.0	0.00	0.0	0.0
3	4	2.011	-2.518	87.99	11.78	5.0	0.0	0.14	0.0	0.0
4	4	0.408	-2.836	35.42	7.90	7.0	0.0	0.00	0.0	0.0

Ejemplo: $H_F \rightarrow \gamma\gamma$

- Momentos transversos $p_T^{\gamma^{1,2}}$ para $M_F = 200$ GeV,



Ejemplo: $H_F \rightarrow \gamma\gamma$

Event Analysis Interface

Particles to Analyze Calculation and Analysis Model Training Significance

1. Load Processed File

Provide the processed file to view information about it: **Cargar base de datos**

Data size: (495687, 12)
 Number of events: 495687
 Number of Signal events: 110595
 Number of Background events: 385092
 Signal fraction: 22.31%

2. Process data

Variables dropped due to high correlation: []

Percentage of NaN values by column:
 - Unnamed: 0: 0.0%
 - Evento: 0.0%
 - M_inv_a1_a2: 0.0%
 - Etab1: 0.0%
 - Etab2: 0.0%
 - Phi1: 0.0%
 - Phi2: 0.0%

3. Model Training

Training ratio (Train/Test e.g., 0.8):
 0.8

Variables used for training:
 Unnamed: 0
 Evento
 M_inv_a1_a2
 Etab1
 Etab2
 Phi1
 Phi2
 Pta1
 Pta2
 DeltaR

Train and Optimize Model

Porcentaje para entrenamiento y prueba y selección de variables.

Info Data Impute with MiceForest

Información de los datos

Eliminar variables

Remove Selected Variables

Spreadsheets data_hf_aa_bys_200Gev_shuffled_xgb.csv - Spreadsheets

Home Insert Page Layout Formulas Data Review View

Paste Cut Copy Format Painter

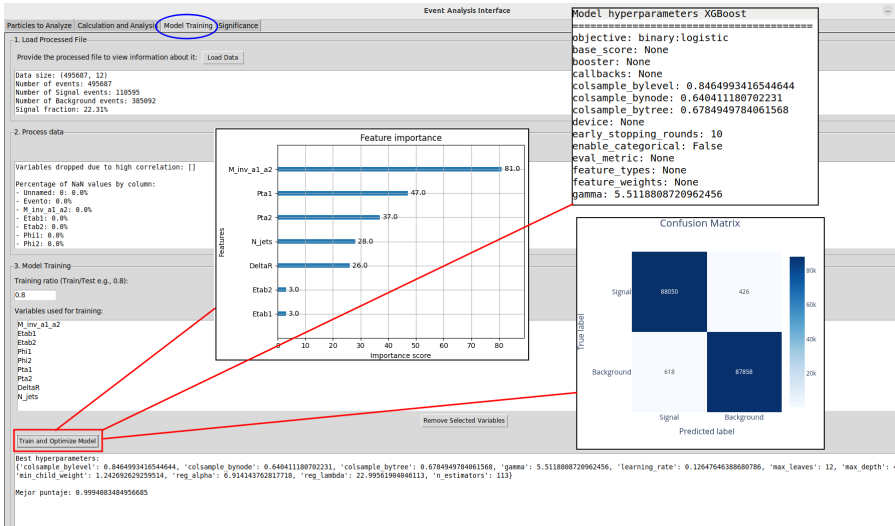
Calibri 11 A A

data_hf_aa_bys_200Gev_shuffled_xgb.csv

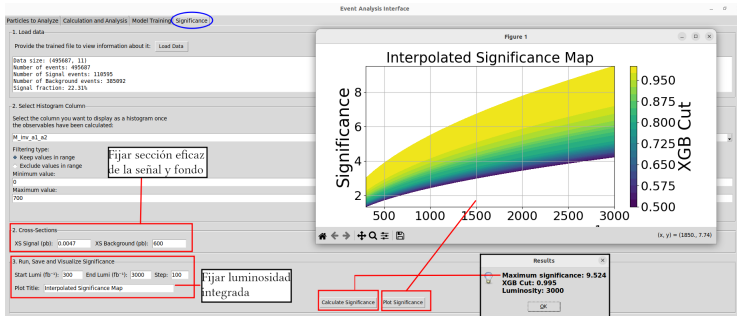
	A	B	C	D	E	F	G	H	I	J
	Unnamed: 0	Evento	M_inv_a1	Etab1	Etab2	Phi1	Phi2	Pta1	Pta2	DeltaR
1	0	1	199.7702	0.397	-0.35	-1.36	2.485	107.49	90.34	3.916
2	0	1	200085	35.35224	-1.071	-0.639	3.06	0.288	19.62	15.36
3	2	374137	25.7884	0.524	0.708	-2.259	0.877	13.36	12.34	3.141
4	2	41576	198.687	1.208	0.719	-0.817	2.448	101.35	92.11	3.301
5	4	437287	32.00093	-1.657	-1.275	-2.822	-0.173	18.29	14.32	2.676
6	5	456347	27.06796	0.412	-0.125	0.248	-2.792	13.85	13.18	3.087
7	6	5055	205.9777	1.468	0.013	3.069	-0.457	82.96	80.99	3.814
8	7	402227	41.42083	0.528	-1.399	-1.132	-1.064	21.01	16.27	1.928
9	8	355646	51.79929	2.255	-0.369	2.096	-0.955	15.54	10.89	4.024
10	8	482760	48.87289	-1.847	-0.319	0.456	-2.679	19.08	18.34	3.487
11	10	297041	68.90449	0.042	-1.585	1.443	-1.909	26.04	25.18	3.725
12	11	186204	67.15941	-0.919	-0.309	0.053	-2.85	36.04	29.58	2.952
13	11	481948	34.82701	0.899	0.877	1.534	-0.715	24.17	12.43	3.228
14	15	13	458717	51.06435	0.609	0.502	1.324	-1.825	26.85	24.21
15	14	325409	22.73548	-2.167	-1.649	-2.064	0.848	11.09	11.04	2.957
16	15	412873	44.52659	0.5	-1.193	2.787	-0.382	18.21	14.29	3.592
17	16	85999	198.6494	0.938	0.244	-1.239	2.168	137.38	64.83	3.477
18	17	44025	208.582	1.072	1.479	-0.936	2.21	107.57	97.02	3.178
19	20	148785	43.69415	0.946	-0.862	1.511	-1.24	16.74	14.06	3.291
20	19	399801	30.121	-1.358	-2.19	-1.368	2.071	14.34	13.62	3.538
21	20	63307	199.9106	-0.98	-0.98	2.469	-1.54	169.42	71.62	4
22	21	139569	87.33125	2.417	2.383	1.722	-1.372	49.68	36.39	3.044
23	22	206489	22.11068	0.848	0.599	0.202	-2.622	12.18	10.13	2.634
24	23	472241	62.38775	1.155	-1.07	0.519	-2.61	18.7	18.32	3.835
25	23	292896	26.88311	-0.117	1.127	0.831	-2.113	12.21	10.35	3.15
26	24	381866	30.12095	-0.592	-0.392	-1.825	1.191	20.74	10.87	3.022
27	25	208213	52.77942	-1.244	-0.486	-1.101	-2.089	20.55	17.32	3.625
28	26	355110	28.77366	1.46	0.768	-1.883	1.434	13.84	13.39	3.385
29	30	288	196.172	0.24951	0.658	-0.669	2.784	-0.537	13.42	10.62
30	31	151105	23.98724	-2.069	-2.316	2.398	-0.532	13.13	10.91	2.940
31	30	150967	30.83175	-1.488	-1.788	1.154	-1.849	15.33	15.23	3.017
32	31	422276	45.1483	-0.289	0.811	-2.094	0.713	19.77	18.94	2.5
33	32	211841	24.17564	0.171	0.644	2.118	-1.861	13.84	11.84	4.007
34	33	325130	29.03701	1.079	0.894	-2.479	0.373	17.52	11.93	3.127
35	34	483372	43.39442	2.029	0.688	-0.662	2.496	18.39	16.83	3.430
36	35	324357	35.65248	-1.819	-1.713	-2.401	0.821	20.57	15.43	3.223
37	36	220669	30.14902	-0.831	0.202	-1.478	-1.778	14.15	12.8	2.943
38	37	44867	198.8414	1.23	2.351	-2.412	0.689	89.99	81.48	3.297
39	38	47785	199.8251	1.113	-0.258	-0.594	-2.393	161.8	58.75	2.185
40	39	138250	137.2603	1.288	1.109	2.505	-0.47	76.18	61.76	2.98
41	40	388039	177.0496	0.616	-0.812	-0.748	2.113	80.21	61.72	3.197
42	41	285419	28.90452	-0.345	-1.126	2.293	-0.603	14.6	12.49	2.995

data_hf_aa_bys_200Gev_shuffled_xgb.csv

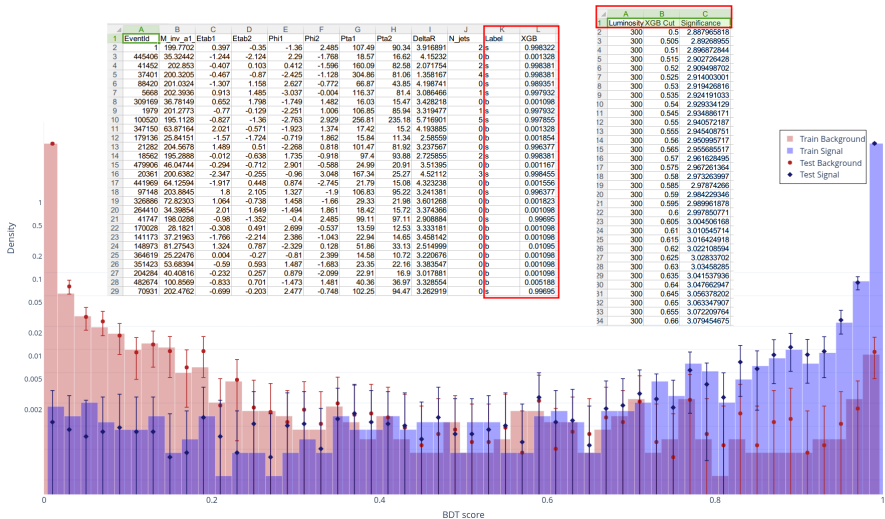
Sum=0 Average=0 Count=12 AutoBackup 100%

$$H_F \rightarrow \gamma\gamma$$


Ejemplo: $H_F \rightarrow \gamma\gamma$



Ejemplo: $H_F \rightarrow \gamma\gamma$



Limitaciones

- ▶ Únicamente admite formato .lhco (por el momento),
- ▶ MAYA sólo hace uso de árboles de decisión boosteados,
- ▶ Dependiendo de los recursos computacionales, la generación de base de datos puede demorar.

Conclusiones y perspectivas

- ▶ Hemos presentado una aplicación programada en Python: MAYA , cuyo objetivo principal es evaluar significancias estadísticas mediante árboles de decisión potenciados (y/o cortes cinemáticos tradicionales),
- ▶ Si el usuario no cuenta con su propia base de datos para el entrenamiento y prueba, MAYA crea una base a partir de archivos a nivel detector .lhco,
- ▶ Se ha creado una base sólida para robustecer a MAYA a través de métodos adicionales de aprendizaje automático,
- ▶ Integrar Redes Neuronales Artificiales,
- ▶ Implementar análisis de cálculos que surgen a nivel de un lazo y materia oscura.

¡Muchas gracias!

