

---

# **Domain- Science & Research**

**Dated: 3<sup>rd</sup> September 2025**

Title of the Project

---

# Particle Physics Event Classification

# Problem Statement:

IN THE FIELD OF PARTICLE PHYSICS, ACCURATE IDENTIFICATION AND CLASSIFICATION OF EVENTS ARE CRUCIAL FOR UNDERSTANDING FUNDAMENTAL PARTICLES AND THEIR INTERACTIONS. THE PROVIDED DATASET CONTAINS VARIOUS FEATURES DERIVED FROM PARTICLE PHYSICS EXPERIMENTS, AND THE GOAL IS TO BUILD A MACHINE-LEARNING MODEL TO CLASSIFY EVENTS INTO SIGNAL (**s**) AND BACKGROUND (**b**) CATEGORIES. THIS CLASSIFICATION AIDS IN DISTINGUISHING EVENTS OF INTEREST (SIGNAL) FROM BACKGROUND NOISE.

# Objective/Goal

---

The primary goal of this machine learning project is to develop a model that can accurately predict whether a given set of experimental features corresponds to a signal or background event. Successful classification contributes to the advancement of particle physics research by automating the identification of events that may indicate the presence of specific particles or phenomena.

# Data Description

- **EventId**: Identifier for each event in the experiment.
- **DER\_mass\_MMC**: Derived mass of the Missing Mass Calculator. Represents the calculated mass of a system using the missing energy in the experiment.
- **DER\_mass\_transverse\_met\_lep**: Indicates the mass calculated from the transverse components of missing energy and lepton.
- **DER\_mass\_vis**: Derived mass of the visible part of the system. Represents the mass of the visible particles in the system, excluding invisible or undetected particles.
- **DER\_pt\_h**: Transverse momentum of the Higgs boson. Indicates the momentum of the Higgs boson in the transverse plane, providing information about its motion.
- **DER\_deltaeta\_jet\_jet**: The absolute difference in pseudorapidity between the two jets.

Describes the angular separation between two jets in the experiment.

- **DER\_mass\_jet\_jet**: Derived mass of the two jets.

- **DER\_lep\_eta centrality**: Pseudorapidity centrality of the lepton concerning jets.
- Pseudorapidity is a spatial coordinate used in particle physics. This feature likely indicates the centrality of the lepton (a charged particle, possibly an electron or a muon) concerning jets in the event. The centrality provides information about the lepton's position relative to the jets.
- **PRI\_tau\_pt**: Transverse momentum of the tau.
- Transverse momentum is the momentum of a particle in the direction perpendicular to the beamline. This feature represents the transverse momentum of the tau particle, which is a heavy, charged particle often involved in processes like tau decays.
- **PRI\_tau\_eta**: Pseudorapidity of the tau.
- Pseudorapidity is a measure of the angle of a particle's trajectory relative to the beamline. This feature indicates the pseudorapidity of the tau particle, providing information about its direction.
- **PRI\_tau\_phi**: Azimuthal angle of the tau.
- Azimuthal angle defines the rotation of a particle's trajectory around the beamline. This feature represents the azimuthal angle of the tau particle, indicating its orientation in the transverse plane.
- **PRI\_lep\_pt**: Transverse momentum of the lepton.
- Similar to PRI\_tau\_pt, this feature represents the transverse momentum of the lepton (electron or muon) in the event.
- **PRI\_lep\_eta**: Pseudorapidity of the lepton.
- Similar to PRI\_tau\_eta, this feature indicates the pseudorapidity of the lepton, providing information about its angular position relative to the beamline.
- **PRI\_lep\_phi**: Azimuthal angle of the lepton.
- Similar to PRI\_tau\_phi, this feature represents the azimuthal angle of the lepton, indicating its orientation in the transverse plane.

- **PRI\_met**: Missing transverse energy.

- Missing transverse energy is a crucial concept in particle physics. It represents the imbalance in transverse energy in an event, suggesting the presence of undetected particles or neutrinos.

- **PRI\_met\_phi**: Azimuthal angle of the missing transverse energy.

- Similar to PRI\_tau\_phi and PRI\_lep\_phi, this feature represents the azimuthal angle of the missing transverse energy, indicating its orientation in the transverse plane.

- **PRI\_met\_sumet**: Sum of the transverse energy of all objects in the event.

- This feature represents the sum of the transverse energy (energy component perpendicular to the beamline) of all objects (particles) detected in the event. It provides a measure of the overall energy flow in the transverse plane.

- **PRI\_jet\_num**: Number of jets in the event.

- Indicates the count of jets observed in the event. Jets are collimated sprays of particles resulting from high-energy interactions. Knowing the number of jets is crucial for understanding the event's characteristics.

- **PRI\_jet\_leading\_pt**: Transverse momentum of the leading jet.

- Represents the transverse momentum of the jet with the highest transverse momentum in the event. The leading jet is the one with the most significant momentum contribution.

- **PRI\_jet\_leading\_eta**: Pseudorapidity of the leading jet.

- Indicates the pseudorapidity (angular position) of the leading jet, providing information about its direction in the detector.

- **PRI\_jet\_leading\_phi**: Azimuthal angle of the leading jet.

- Represents the azimuthal angle of the leading jet, indicating its orientation in the transverse plane.

- **PRI\_jet\_subleading\_pt**: Transverse momentum of the subleading jet.
  - Represents the transverse momentum of the second-highest transverse momentum jet in the event (subleading jet).
- **PRI\_jet\_subleading\_eta**: Pseudorapidity of the subleading jet.
  - Indicates the pseudorapidity of the subleading jet, providing information about its angular position in the detector.
- **PRI\_jet\_subleading\_phi**: Azimuthal angle of the subleading jet.
  - Represents the azimuthal angle of the subleading jet, indicating its orientation in the transverse plane.
- **PRI\_jet\_all\_pt**: Sum of the transverse momentum of all jets in the event.
  - This feature represents the sum of the transverse momentum of all jets detected in the event. It provides a measure of the overall transverse momentum contributed by all jets.
- **Weight**: A weight associated with each event.
  - The weight is a numerical value assigned to each event. It is often used as a scaling factor to account for the significance or importance of each event in the analysis. The weight may be applied during training and evaluation to adjust the contribution of each event to the model.
- **Target column(Label)**: The target variable, indicating the class or label of each event (**s** or **b**). This is the target variable where each event is labeled as either signal (**s**) or background (**b**).



# Dataset structure

---

No of Columns- 33

No of records/rows- 250001

# Dataset sample

EventId	DER_mas s_MMC	DER_mas s_transver se_met_le p	DER_mas s_vis	DER_pt_h	DER_delta eta_jet_jet	DER_mas s_jet_jet	DER_prod eta_jet_jet	DER_delta r_tau_lep_t	DER_pt_to pt	DER_sum pt	DER_pt_r atio_lep_taphi_centra	DER_met_ eta_centra	DER_lep_ lity	PRI_tau_p t
100000	138.47	51.655	97.827	27.98	0.91	124.711	2.666	3.064	41.928	197.76	1.582	1.396	0.2	32.638
100001	160.937	68.768	103.235	48.146	-999	-999	-999	3.473	2.078	125.157	0.879	1.414	-999	42.014
100002	-999	162.172	125.953	35.635	-999	-999	-999	3.148	9.336	197.814	3.776	1.414	-999	32.154
100003	143.905	81.417	80.943	0.414	-999	-999	-999	3.31	0.414	75.968	2.354	-1.285	-999	22.647
100004	175.864	16.915	134.805	16.405	-999	-999	-999	3.891	16.405	57.983	1.056	-1.385	-999	28.209
100005	89.744	13.55	59.149	116.344	2.636	284.584	-0.54	1.362	61.619	278.876	0.588	0.479	0.975	53.651
100006	148.754	28.862	107.782	106.13	0.733	158.359	0.113	2.941	2.545	305.967	3.371	1.393	0.791	28.85
100007	154.916	10.418	94.714	29.169	-999	-999	-999	2.897	1.526	138.178	0.365	-1.305	-999	78.8
100008	105.594	50.559	100.989	4.288	-999	-999	-999	2.904	4.288	65.333	0.675	-1.366	-999	39.008
100009	128.053	88.941	69.272	193.392	-999	-999	-999	1.609	28.859	255.123	0.599	0.538	-999	54.646
100010	-999	86.24	79.692	27.201	-999	-999	-999	2.338	27.201	81.734	1.75	-1.412	-999	29.718
100011	114.744	10.286	75.712	30.816	2.563	252.599	-1.401	2.888	36.745	239.804	1.061	1.364	0.769	35.976
100012	145.297	64.234	103.565	106.999	-999	-999	-999	2.183	24.66	192.245	0.576	0.689	-999	62.89
100013	82.488	31.663	64.128	8.232	-999	-999	-999	2.823	8.232	58.649	1.303	-1.414	-999	25.47
100014	-999	109.412	14.398	17.323	-999	-999	-999	0.472	17.323	62.565	1.774	-0.272	-999	22.552
100015	111.026	32.096	75.271	23.067	-999	-999	-999	3.205	23.067	69.649	1.276	-1.414	-999	30.606
100016	114.256	4.351	67.963	47.221	-999	-999	-999	2.954	26.243	100.93	1.145	0.218	-999	30.145
100017	127.861	50.953	77.267	26.967	-999	-999	-999	2.833	26.967	79.503	1.586	1.401	-999	30.739
100018	-999	85.186	68.827	5.042	-999	-999	-999	2.116	5.042	71.443	1.558	-1.351	-999	27.931
100019	-999	88.767	115.058	15.337	-999	-999	-999	2.879	15.337	58.211	0.875	-1.395	-999	31.046
100020	-999	89.705	41.765	18.437	-999	-999	-999	1.395	18.437	57.157	1.082	-1.178	-999	27.453
100021	90.736	18.674	60.231	25.156	-999	-999	-999	2.363	25.156	64.833	0.749	-0.8	-999	37.06
100022	87.075	38.217	67.041	2.347	-999	-999	-999	2.852	2.347	65.281	1.276	-1.409	-999	28.688
100023	141.481	0.736	111.581	174.075	1.955	364.344	-0.923	1.335	6.663	440.859	0.652	1.042	0.207	98.565

PRI_tau_eta	PRI_tau_phi	PRI_lep_pt	PRI_lep_eta	PRI_lep_phi	PRI_met	PRI_met_phi	PRI_met_su met	PRI_jet_num	PRI_jet_leadi ng_pt	PRI_jet_leadi ng_eta	PRI_jet_leadi ng_phi	PRI_jet_subl eading_pt	PRI_jet_subl eading_eta	PRI_jet_subleading_phi	PRI_jet_all_p t	Weight	Label
1.017	0.381	51.626	2.273	-2.414	16.824	-0.277	258.733	2	67.435	2.15	0.444	46.062	1.24	-2.475	113.497	0.002653	s
2.039	-3.011	36.918	0.501	0.103	44.704	-1.916	164.546	1	46.226	0.725	1.158	-999	-999	-999	46.226	2.233584	b
-0.705	-2.093	121.409	-0.953	1.052	54.283	-2.186	260.414	1	44.251	2.053	-2.028	-999	-999	-999	44.251	2.347389	b
-1.655	0.01	53.321	-0.522	-3.1	31.082	0.06	86.062	0	-999	-999	-999	-999	-999	-999	0	5.446378	b
-2.197	-2.231	29.774	0.798	1.569	2.723	-0.871	53.131	0	-999	-999	-999	-999	-999	-999	0	6.245333	b
0.371	1.329	31.565	-0.884	1.857	40.735	2.237	282.849	3	90.547	-2.412	-0.653	56.165	0.224	3.106	193.66	0.083414	b
1.113	2.409	97.24	0.675	-0.966	38.421	-1.443	294.074	2	123.01	0.864	1.45	56.867	0.131	-2.767	179.877	0.002653	s
0.654	1.547	28.74	0.506	-1.347	22.275	-1.761	187.299	1	30.638	-0.715	-1.724	-999	-999	-999	30.638	0.018636	s
2.433	-2.532	26.325	0.21	1.884	37.791	0.024	129.804	0	-999	-999	-999	-999	-999	-999	0	5.296003	b
-1.533	0.416	32.742	-0.317	-0.636	132.678	0.845	294.741	1	167.735	-2.767	-2.514	-999	-999	-999	167.735	0.001502	s
-0.866	2.878	52.016	0.126	-1.288	51.276	0.688	250.178	0	-999	-999	-999	-999	-999	-999	0	2.299504	b
-0.669	-0.342	38.188	-0.165	2.502	22.385	2.148	290.547	3	76.773	-0.79	0.303	56.876	1.773	-2.079	165.64	0.30717	b
-0.766	-1.632	36.237	0.722	-0.035	43.91	-1.907	232.362	1	93.117	-0.97	1.943	-999	-999	-999	93.117	1.681611	b
-0.654	-2.99	33.179	-1.665	-0.354	12.439	1.433	163.42	0	-999	-999	-999	-999	-999	-999	0	2.183892	b
1.389	1.34	40.013	1.856	1.412	75.197	-1.583	198.616	0	-999	-999	-999	-999	-999	-999	0	2.151199	b
-1.107	-1.903	39.043	-1.944	1.191	19.959	2.415	122.176	0	-999	-999	-999	-999	-999	-999	0	0.018636	s
0.484	-0.929	34.522	-0.215	1.941	41.899	2.055	191.568	1	36.263	-0.766	-0.686	-999	-999	-999	36.263	1.454848	b
-0.635	2.603	48.764	-0.343	-0.862	17.557	-2.975	211.72	0	-999	-999	-999	-999	-999	-999	0	0.001503	s
1.175	2.356	43.512	2.332	0.584	44.698	-2.033	151.816	0	-999	-999	-999	-999	-999	-999	0	5.121624	b
1.38	0.451	27.165	-1.486	0.724	72.981	-2.577	115.145	0	-999	-999	-999	-999	-999	-999	0	5.979351	b
1.58	2.51	29.704	0.341	1.869	67.909	-1.169	225.139	0	-999	-999	-999	-999	-999	-999	0	2.198594	b
1.537	-2.616	27.773	1.161	1.335	27.431	0.645	169.061	0	-999	-999	-999	-999	-999	-999	0	1.681611	b
1.739	-2.975	36.594	2.367	-0.193	12.556	2.008	68.527	0	-999	-999	-999	-999	-999	-999	0	4.569368	b
0.19	-1.506	64.285	1.405	-0.952	17.96	-0.973	454.785	2	195.533	1.156	1.416	82.477	-0.798	-2.785	278.009	0.001503	s

# Instructions

---

The dataset will be given through a drive link in your project group on the Learnbay app

You have to submit the project with a Ppt presentation by **Saturday 13th September 2025 by EOD.**

Kindly submit your '**XYZ.ipynb**' file and '**XYZ.ppt**' to **interviewprep@learnbay.co** within the timeframe, submission of the project after the due date will be considered disqualified. Late submissions will be considered for a valid reason.

After submission of the project you'll get a link to book a time for the project presentation.

# Selection of feedback

---

Selection of candidates will be based on their **approach to building a model, presentation skills(Storytelling skills), and subject knowledge points(a mock round)(Questions related to ML and Python for Data Science**

**Note: you need to score 80% to clear this round.**

Once the presentation is done every candidate will get their feedback during the session and outcome and score via mail with the status of whether they are selected or not.

Selected candidates' data will be shared with the placement team for 1 on 1 resume session.

Candidates not selected in this process will be carried forward to the next project.

Kindly do not book multiple slots, if found it shall considered as cancelled. If any change in the slot date and time kindly inform or cancel the previous slot.

If you are absent or unable to present on the day of the presentation, getting another will be subject to availability.