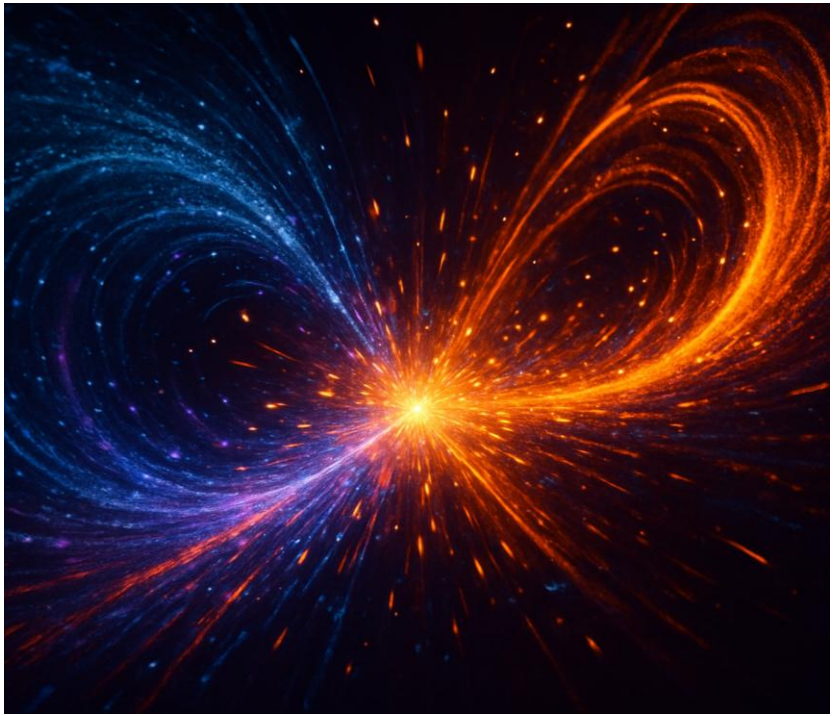
Particle Physics Event Classification

Distinguish rare Higgs-boson signal events from overwhelming background noise in proton–proton collisions.
Build a robust machine-learning pipeline to separate the signal (s) from background (b) using rich kinematic features derived from detector data.



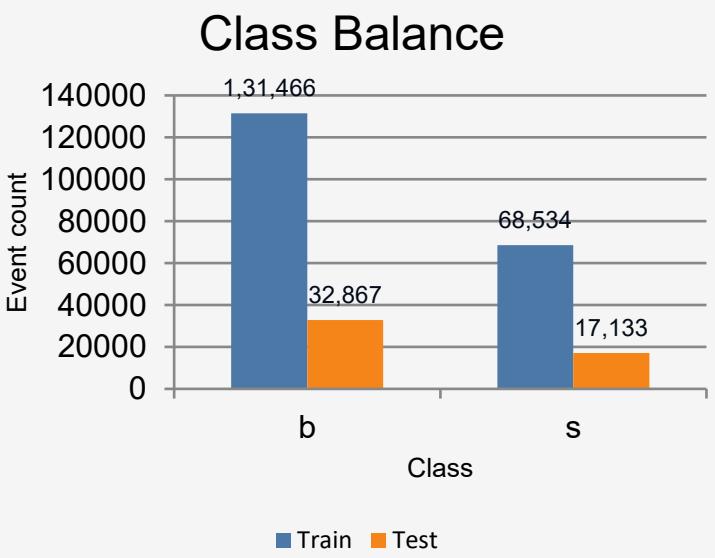
13 Sept 2025

[1]

Methodology

- Preprocessing: impute missing values, cap outliers and drop irrelevant columns.
- Base models: train XGBoost, LightGBM & CatBoost with tuned hyper-parameters.
- Stacking: use base probabilities as features and fit a logistic-regression meta-learner.
- Cross-validation & threshold tuning (0.49) balance recall versus precision.

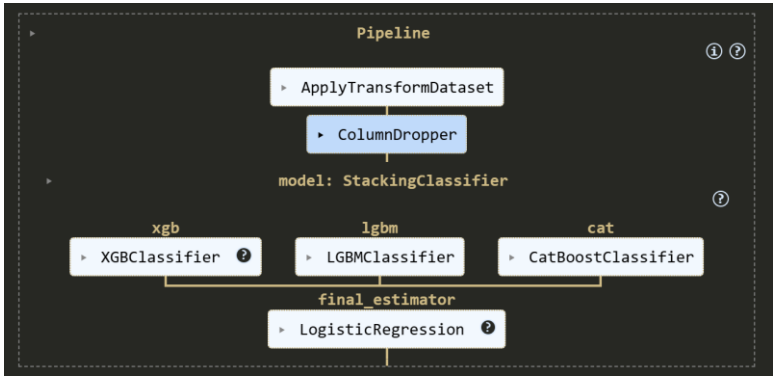
Statistic	Value
Total events (n)	200,000
Signal events (s)	68,534
Background events (b)	131,466
Raw features	30
Features used	18



[1]

Model Architecture

- Transform dataset: impute missing values and cap outliers (ApplyTransformDataset).
- Drop non-feature columns (ColumnDropper).
- Base estimators: XGBoost, LightGBM & CatBoost trained on cleaned features.
- Stacking: collect base probabilities and feed them into a logistic-regression meta-learner.
- Cross-validated predictions prevent overfitting and enable robust ensemble learning.

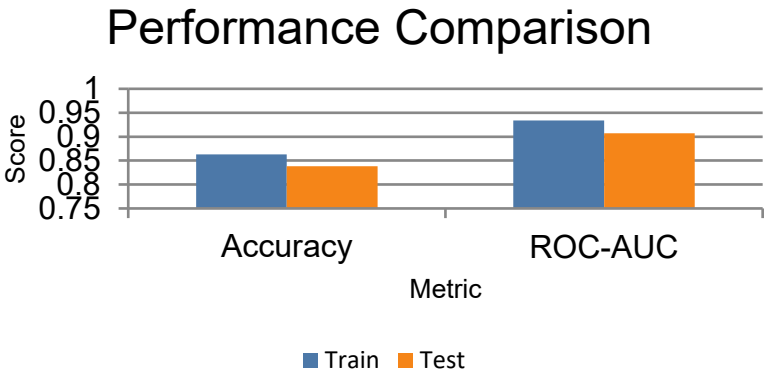


Model	n_estimators	max_depth	learning_rate
XGBoost	700	7	0.1
LightGBM	700	7	0.1
CatBoost	700	7	0.1

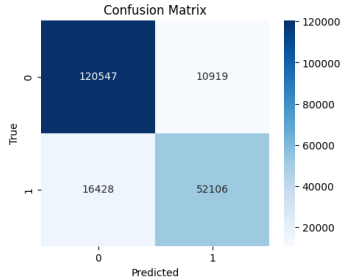
[2]

Results & Summary

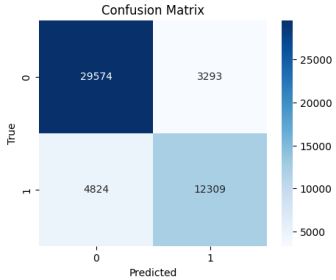
- High training AUC (~0.93) and accuracy (~0.86) show the ensemble captures signal–background patterns well.
- Cross-validated performance remains strong (AUC ≈ 0.91, ACC ≈ 0.84) on unseen folds and the held-out test set.
- The stacking ensemble combines XGBoost, LightGBM and CatBoost under a logistic meta-learner.
- More false negatives in training hint at mild overfitting; calibration and regularisation could help.
- Future work: investigate deeper architectures and fine-tune thresholds for better recall.



Train



Test



Dataset	Accuracy (μ ± σ)	ROC-AUC (μ ± σ)
Train	0.863	0.934
CV (mean ± σ)	0.838 ± 0.002	0.908 ± 0.002
Test	0.838	0.907

[1] [2]