



Contribution ID: 181

Type: Poster

Running Experience with Optuna for the Extraction of a HEP Signal by XGBoost

Wednesday 10 September 2025 11:00 (30 minutes)

Hyperparameter optimization plays a crucial role in achieving high performance and robustness for machine learning models, such those used in complex classification tasks in High Energy Physics (HEP).

In this study, we investigate and experience the usage of Optuna, a rather new, modern and scalable optimization tool in the framework of a realistic signal-versus-background classification scenario carried out by applying XGBoost on CMS Open Data.

The chosen classification task consists in extracting the signal associated to the decay mode of $B_s \rightarrow J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ by means of a gradient boost tree (XGBoost) trained on both Monte Carlo simulated signal sample and a background one taken from the data as invariant mass sidebands, in the $\phi(\rightarrow K^+K^-)$ spectrum. The optimization process of XGBoost is guided by Optuna with the aim to maximize the area under the ROC curve (AUC) while applying an overfitting control mechanism, whereas the Punzi Figure of Merit is used for a performant extraction of the signal within XGBoost.

This work demonstrates how Optuna is a suitable tool that enables efficient and effective exploration of the hyperparameter space in commonly used HEP workflows, while providing valuable diagnostics on the automated model optimization.

Significance

This contribution discusses the use of Optuna, a modern hyperparameter optimization framework, within a realistic HEP classification workflow (using CMS Open Data). Although commonly adopted in other data science domains, Optuna is not widely used in HEP yet. We demonstrate that its integration with XGBoost allows an efficient and effective performance when applied to a typical signal-versus-background task to extract the physical signal, also highlighting its efficiency, flexibility, and diagnostic capabilities.

References

Experiment context, if any

Authors: DI FLORIO, Adriano (CC-IN2P3); POMPILI, Alexis (Universita e INFN, Bari (IT)); SOZBILIR, Umit (Universita e INFN, Bari (IT)); MASTRAPASQUA, Vincenzo (Universita e INFN, Bari (IT))

Presenter: SOZBILIR, Umit (Universita e INFN, Bari (IT))

Session Classification: Poster session with coffee break

