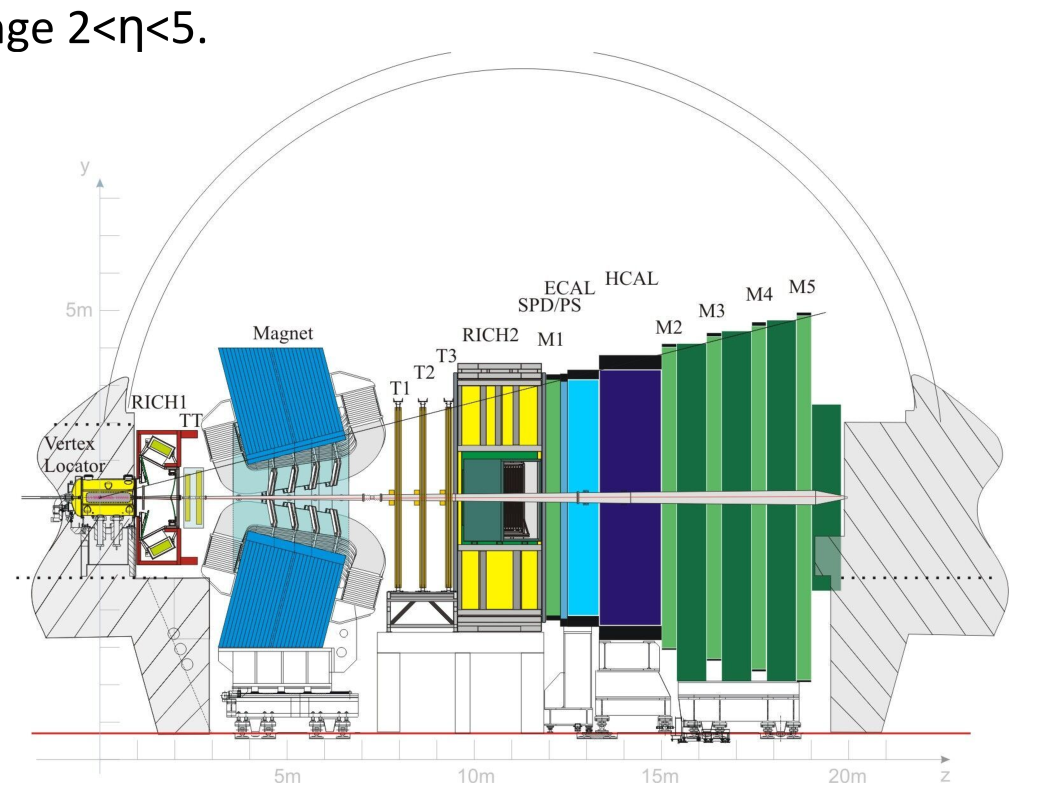
Machine learning based long-lived particle reconstruction algorithm for the LHCb experiment

Adam Dendek[1], Tomasz Szumlak[1]
[1]AGH University of Science and Technology

LHCb detector

The **LHCb detector** is a single-arm forward spectrometer at the LHC with a pseudorapidity in the range $2<\eta<5$.



LHCb track types

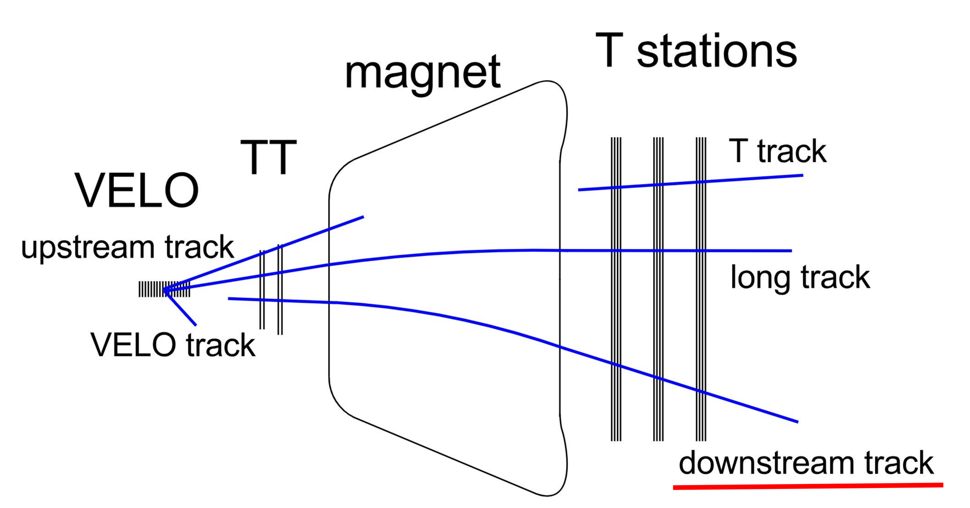
The tracks reconstructed in the LHCb detector are divided into types depending on the sub detectors in which they are reconstructed.

Long Tracks:

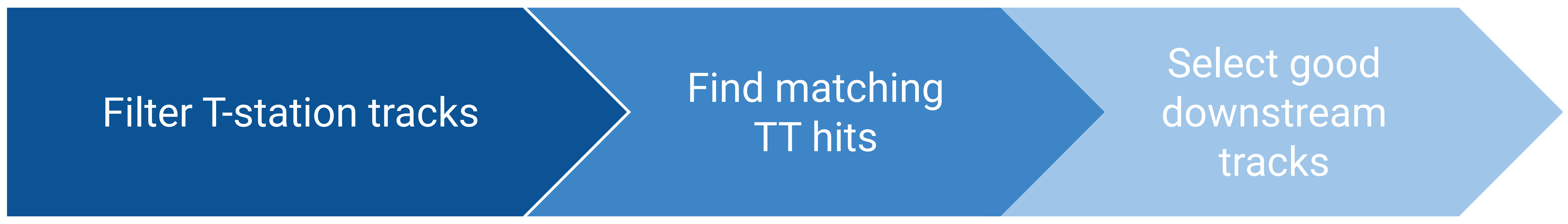
- Hits at least in VELO and T stations
- Excellent momentum resolution
- Used in majority of analyses

Downstream Tracks:

- Hits in TT and T stations.
- Daughters of long lived particles
- Example: $\mathbf{\Lambda^0 \to p^+ + \pi^-}$



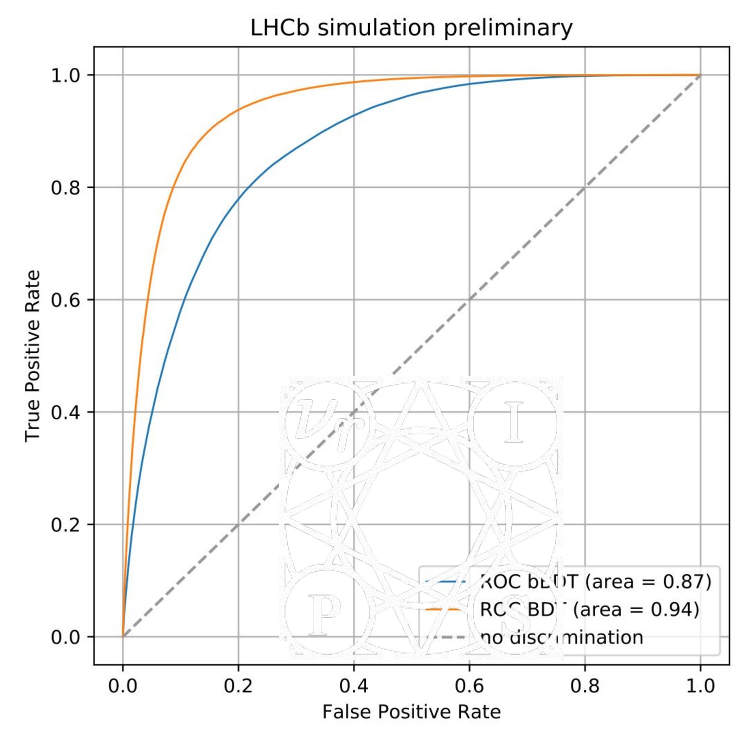
Downstream tracking sequence



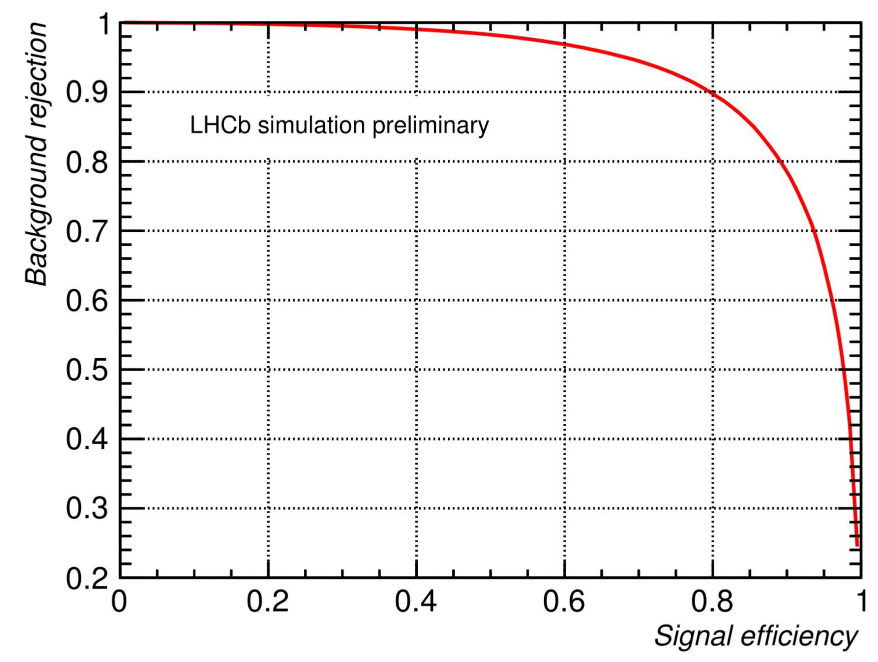
Machine learning in the LHCb downstream tracking algorithm

Downstream tracking contains two Machine Learning classifiers.

- The first is designed to **reject as much fake T tracks as possible**:
  - Implemented as *bonsai Boosted Decision Trees (bBDT)*,
  - Discretized 11D feature space,
  - Inputs: T tracks topological variables, p, pt, …
  - Operating point: bBDT removes 30% of fake T tracks,
  - **Fast evaluation time O(1).**
- The second for the final selection of downstream tracks:
  - further reducing fake tracks implemented as *Neural Network*
  - Inputs: p, p_t, track position, number of layers in TT, …
- **Improved fake tracks reduction and signal efficiency gain: O(3-5)%**



ROC curve for selection true T-seed tracks using a simulated $B \to J/\Psi K_s^0$ sample. The orange curve corresponds to the performance of the classifier before binarization and the blue corresponds to binarized version.



ROC curve for final accepting tracks using a sample of $\mathbf{D^{*+} \to D^0 \pi^+ (D^0 \to K_s^0 K^+ K^-)}$ decays.

Interpretability of the machine learning classifier

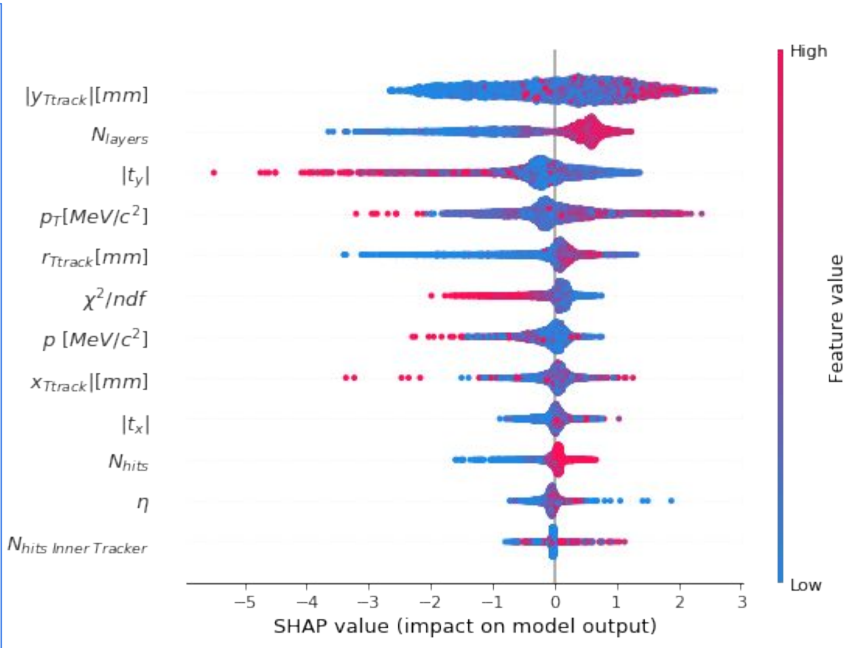
One of the crucial problems when building a complicated machine learning model is a lack of interpretability of its prediction. This fact raises the question of why the researcher should trust the model. To make sure that the model provides reliable predictions two methods, were proposed: **SHAP** (Shapley Additive exPlanations) and **LIME** (Local Interpretable Model-Agnostic Explanations).

SHAP method:

- SHAP connects **game theory** with local explanations.
- In a game there is N players, each of it may cooperates and obtains certain gain from cooperation.
- Taking into consideration that some players may contribute more to the coalition the question is how to fairly share the payoffs?
- The answer for this question is known as **Shapley value:**

$$\varphi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(N-|S|-1)!}{N!}(v(S\cup\{i\}) - v(S))$$

V value of coalition **S**

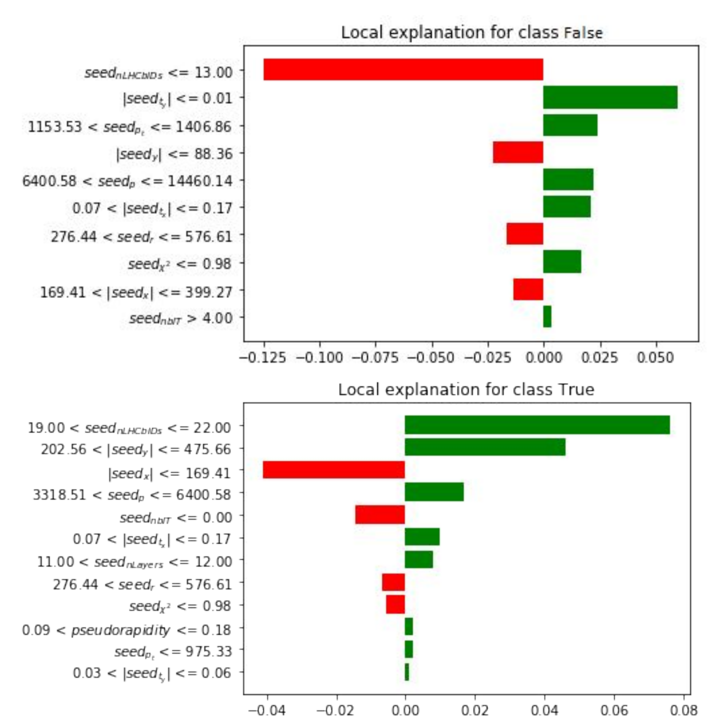


LIME method:

- The idea is to simplify complex model prediction by its **interpretable representation** (linear model).
- Conceptually very close to **Taylor series** approximations
- The interpretation is calculated for a **single instance**
- In order to ensure both local **fidelity** and **interpretability**, the explanation needs to minimize the following formula:

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}}\ \boxed{\mathcal{L}(f, g, \pi_x)} + \boxed{\Omega(g)}$$

π similarity kernel,
G class of human interpretable models (e.g. linear model, decision trees)



Conclusions

The new algorithm to reconstruct tracks of daughters of long-lived particles in LHCb considerably improves on the performance in Run II data taking conditions over the previous algorithm in terms of track reconstruction efficiency, fake track rate, and also CPU time consumption is reduced by about one third. Within this study, two interpretability methods of the machine learning model were proposed, which allowed increasing the trust in the classifier predictions.

Third Workshop on Machine Learning and the Physical Sciences
Neural Information Processing Systems (NeurIPS)

Contact

Adam Dendek
Email: adendek@agh.edu.pl

References

1. The LHCb Collaboration et al 2008 JINST 3 S08005
2. V V Gligorov and M Williams, Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree, J. Instrum. 02(2013)
3. A. Davis, A. Dendek, et al. PatLongLivedTracking: a tracking algorithm for the reconstruction of the daughters of long-lived particles in LHCb, LHCb public note in preparation
4. S. Lundberg, S. Lee, A Unified Approach to Interpreting Model Predictions, CoRR abs/1705.07874
5. M. Tulio Ribeiro, S. Singh, C. Guestrin "Why Should I Trust You?": Explaining the Predictions of Any Classifier, [doi:10.1145/2939672.2939778]