



Contribution ID: 118

Type: **Flashtalk with Poster**

Validating Explainable AI Techniques through High Energy Physics Data

Tuesday, 30 April 2024 17:16 (3 minutes)

Within the Compact Muon Solenoid (CMS) Collaboration, various Deep Neural Networks (DNNs) and Machine Learning (MLs) approaches have been employed to investigate the production of a new massive particle that undergoes decay into Higgs Boson pairs (HH) which further decay into a pair of b-quarks and a pair of tau leptons and discriminate the HH signal from the backgrounds.

However, these models are often complex and considered black boxes, making it challenging to interpret how the task was performed and the data analysis review process.

This work aimed therefore to provide a better understanding of how the model is working by validating an established Explainable Artificial Intelligence (AI) technique such as SHapley Additive exPlanations (SHAP), aiming for more interpretable, trustworthy models and predictions.

A data pre-processing pipeline was established to select the most important features to be used as input to the model. First highly correlated data were removed, secondly a feature selection was performed in repeated 5-fold cross-validation based on SHAP values by means of Recursive Feature Elimination (RFE) algorithm. Finally, a fine tuning of the hyperparameters of a gradient boosting algorithm (XGBoost), trained on the SHAP selected features, was done. This XGBoost model was then used to perform the classification task described at the beginning of this abstract. The selected features were later compared with those obtained from a Principal Component Analysis (PCA) performed on the original entire dataset (prior to any pre-processing steps). PCA achieves dimensionality reduction and, therefore, can be thought of as a clustering method to visualize separation of the observations belonging to the two classes (i.e., signal, background) along the principal components. Therefore, the linear combination of the features along the principal components enabled the validation and interpretation of SHAP.

The results obtained with SHAP and PCA agreed on the importance of some of the features used in the classification task. The combination of the two techniques confirmed the reliability of SHAP as an established tool, but also the potential of High Energy Physics (HEP) domain as a new technical validation tool thanks to its high-quality tabular data and well understood underlying causal theory, which might also be exploited into other fields.

Primary authors: MONTELEONE, Mariagrazia (Politecnico di Milano); CAMAGNI, Francesca (Politecnico di Milano); Dr GENNAI, Simone (INFN sezione di Milano – Bicocca, Milano, Italy); Prof. GOVONI, Pietro (Università degli Studi di Milano Bicocca Milano, Italy INFN sezione di Milano – Bicocca Milano, Italy); Prof. PAGANELLI, Chiara (Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano)

Presenter: MONTELEONE, Mariagrazia (Politecnico di Milano)

Session Classification: 3.2 Physics-informed AI & Integration of physics and ML

Track Classification: Session B