



Contribution ID: 64

Type: **Presentation**

Model interpretability methods for high energy physics analysis

Wednesday 6 July 2022 18:35 (15 minutes)

Most modern machine learning models are known as black-box models. By default, these predictors don't provide an explanation as to why a certain event or example has been assigned a particular class or value. Model explainability methods aim to interpret the decision-making process of a black-box model and present it in a way that is easy for researchers to understand. These methods can provide *local* (figuring why a specific input has been assigned a specific output) an *global* (uncovering general dependencies between input features and the output of the model) explanations. In this talk we will cover several popular model-agnostic explainability methods and compare them in explaining the output of a neural network in the scope of high-energy physics analysis. We will also use a modern high accuracy glass-box machine learning model (Explainable Boosting Machine) and show how its predictions can be used to better understand the data.

Agreement to place

Participants agree to post their abstracts and presentations online at the workshop website. All materials will be placed in the form in which they were provided by the authors

Author: ZABORENKO, Andrei (Moscow State University)

Co-authors: DUDKO, Lev (SINP MSU); VOLKOV, Petr; VOROTNIKOV, Georgy; ABASOV, Emil

Presenter: ZABORENKO, Andrei (Moscow State University)

Session Classification: Session 1. ML in Particle Astrophysics and High Energy Physics

Track Classification: Track 1. Machine Learning in Particle Astrophysics and High Energy Physics