



Contribution ID: 31

Type: **not specified**

ShapKit: a Python module dedicated to local explanation of machine learning models

Tuesday, 14 September 2021 16:45 (20 minutes)

Machine Learning is enjoying an increasing success in many applications: defense, cyber security, etc. However, models are often very complex. This is problematic, especially for critical systems, because end-users need to fully understand the decisions of an algorithm (e.g. why an alert has been triggered or why a person has a high probability of cancer recurrence). One solution is to offer an interpretation for each individual prediction based on attribute relevance. Shapley Values, coming from cooperative game theory, allow to distribute fairly contributions for each attribute in order to understand the difference between a predicted value for an observation and a base value (e.g. the average prediction of a reference population). While these values have many advantages, including their theoretical guarantees, they have a strong drawback: the complexity increases exponentially with the number of features. In this talk, we will present and demonstrate ShapKit, a Python module developed by Thales and available in Open Source dedicated to Shapley Values computation in an efficient way for local explanation of machine learning model. We will apply ShapKit on a cybersecurity use case.

Keywords

Interpretability, Local explanation, Shapley Values.

Special/invited session

special SFDS session

Primary authors: THOUVENOT, Vincent; Mr GRAH, Simon (OCTO Technology)

Presenter: THOUVENOT, Vincent

Session Classification: Machine learning and industrial applications (SFdS)