

Laboratorium No.3 (grupy poniedziałkowe)

Informacje wstępne – co przygotować

1. Proszę zainstalować na swoich komputerach ostatnią wersję oprogramowania ORANGE™ ze strony <https://orangedatamining.com/download/#windows>. Po zainstalowaniu podstawowej wersji proszę doinstalować dodatek **Options/Add-ones Explain**.
2. Zaznajomić się z ogólną strukturą oprogramowania. Przyjrzeć się widget'om oraz połączeniom (zbiór i opis widgetów: <https://orangedatamining.com/widget-catalog/>).
3. Zrobić następujące ćwiczenie wstępne: Pobrać zbiór danych (zakładka **Data**, widget **Data set**), (ogłębnić zbiór używając widgetu **Data table**), ewentualnie podprawić: **Select column** oraz znormalizować **Preprocess**. **Uwaga na normalizację bo zmienia wielkości cech**), przeczytać coś o tym zbiorze danych i atrybutach tego zbioru, podglądnąć zbiór (np. **Unsupervised/t-SNE**). Zastosować jakiś klasyfikator z zakładki **Model** potem dokonać ewaluacji modelu przy pomocy widgetu **Test and score**. Dodałem zbiór z przykładem: *basic_examp1.ows* rekomenduję użyć zbiór z UCI Datasets, *Glass*.
4. Przeanalizować przykład zastosowania metod interpretacji SHAP i LIME na przykładzie z: <https://orangedatamining.com/blog/2021/2021-02-10-explaining-models/>.
5. Proszę przygotować sobie schemat Orange dzielący dowolny zbiór na „znany” i przykłady „nieznane do interpretacji”. W przykładzie 4 takie dwa zbiory zostały już spreparowane (`_train` i `_predict`). Orange ma do nich dostęp poprzez widget **Datasets**. Na zajęciach otrzymają Państwo jeden zbiór np. *Titanic*, dlatego trzeba mieć gotowy schemat do jego podziału na treningowy i przykłady. Wyselekcjonowane przykłady nie powinny znajdować się w zbiorze treningowym.
6. Na zajęciach dostaną Państwo podobne zadania, tylko z innymi zbiorami. Będziemy badać jak wyniki interpretowalności zależą od modelu, wybranych przykładów, a także sposobu preprocessingu danych.
7. Osoby nieprzygotowane do zajęć (brak instalacji oprogramowania, brak wiedzy jego podstawowego użycia, brak gotowego schematu z punktu 4) nie będą do nich dopuszczone.