Laboratorium No.1 (wszystkie grupy)

Informacje wstępne – co przygotować

1. Proszę na zajęciach używać swoich laptopów lub wykonać poniższe czynności na swoich kontach lab. Proszę:
   1. zainstalować ostatnią wersję oprogramowania ORANGETM ze strony <https://orangedatamining.com/download/#windows>. Po zainstalowaniu podstawowej wersji proszę doinstalować dodatek **Options/Add-ones Explain**. Orange to szybki no-programming tool wykonany w całości przy pomocy znanych repozytoriów Pythonia (Scikit-learn, NumPy, Matplotlib etc.).
   2. Oprogramowanie OrangeTM można używać w realizacji zadań lab (jest szybsze w projektowaniu prototypów prostych systemów ML), ale generalnie proszę traktować to raczej jak ciekawostkę służącą do celów dydaktycznych! Załączam do zadania przykład użycia (basic\_exmpl.ows)
   3. Zaznajomić się z ogólną strukturą oprogramowania. Przyjrzeć się widget’om oraz połączeniom (zbiór i opis widgetów: <https://orangedatamining.com/widget-catalog/>).
2. Jeżeli komuś łatwiej i szybciej wykonać to zadanie w notatniku, może to zrobić wykorzystując np. Pythonowe repozytorium SHAP:

[GitHub - shap/shap: A game theoretic approach to explain the output of any machine learning model.](https://github.com/shap/shap) Nie wiem jednak czy wtedy Państwo będą w stanie skończyć ćwiczenie w trakcie trwania lab!

1. Zrobić następujące ćwiczenie wstępne: Pobrać wybrany zbiór danych (w przypadku Orange zakładka **Data**, widget **Data set**), oglądnąć zbiór (używając widgetu **Data table**), ewentualnie podprawić: (Orange: **Select column** oraz znormalizować **Preprocess)**, przeczytać coś o tym zbiorze danych i atrybutach tego zbioru, podglądnąć zbiór przy pomocy wizualizacji t-SNE (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\_exampl.ows* rekomenduję użyć zbiór z UCI Datasets, Wine lub *Glass.*
2. Przeanalizować przykład zastosowania metod interpretacji SHAP i LIme na przykładzie z: <https://orangedatamining.com/blog/2021/2021-02-10-explaining-models/>.
3. Proszę przygotować sobie schemat Orange dzielący dowolny zbiór na „znany” i przykłady „nieznane do interpretacji”. Na zajęciach otrzymają Państwo jeden zbiór*,* dlatego trzeba mieć gotowy schemat do jego podziału na treningowy i przykłady. Nie jest to specjalnie wymagane, bo w widge’cie **Data\_table** wystarczy dokonać selekcji danych które, będą przykładami, a które będą zbiorem treningowym. Wyselekcjonowane przykłady nie powinny znajdować się w zbiorze treningowym.
4. 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.
5. Osoby nieprzygotowane do zajęć (brak instalacji oprogramowania, brak wiedzy jego podstawowego użycia, brak gotowego schematu z punktu 4) nie będą do nich dopuszczone.