

Serwis reprezentatywności

W ramach zadania należy zaimplementować serwis REST'owy w języku Python, który pozwoli na odpytywanie modelu o tak zwaną reprezentatywność obiektu (szczegóły w dalszej części polecenia). Serwis ten powinien udostępniać tylko jeden endpoint - endpoint do predykcji wytrenowanym modelem na przesłanych obiektach. Dla zadanej listy obiektów, REST powinien zwrócić listę reprezentatywności. Gdzie kolejne reprezentatywności są wartościami zwracanymi przez wytrenowany model dla odpowiadających im obiektów.

Miarę reprezentatywności obiektu rozumiemy jako liczbę, która opisuje to jak dobrze dany obiekt reprezentuje inne obiekty w danym zbiorze. Im większa jest reprezentatywność danego obiektu tym więcej w opisywanej przestrzeni danych jest obiektów podobnych.

Istnieje wiele sposobów na wyliczanie reprezentatywności. W tym zadaniu należy zaimplementować tylko jeden z nich, który przebiega następująco:

1. Dzielimy zbiór danych na L mniej więcej równych części w sposób losowy.
2. Dla każdej części zbioru:
 - a. Liczymy odległość każdego obiektu z danego podzbioru od K najbliższych obiektów w danym fragmencie zbioru.
 - b. Obliczamy średnią z tych k wielkości.
 - c. Reprezentatywność obiektu definiujemy jako:

$$\frac{1}{1 + \text{średnia_odległość_k_sąsiadów}}$$

- d. Gdy mamy mapowanie obiekt \rightarrow reprezentatywność, uczymy model, który będzie w stanie przewidywać reprezentatywność na nowych danych.
3. Zbieramy wszystkie L wyliczonych modeli, nasz finalny model będzie ich ensemble'm zwracającym średnią z ich predykcji.

Do wytrenowania modelu można skorzystać z dowolnego zbioru danych o otwartej licencji. W zadaniu nie trzeba implementować endpointu do trenowania modelu. Wytrenowany wcześniej model powinien być dostępny wewnątrz serwisu.

Rozwiązanie należy udostępnić w repozytorium GitHub.