

Opis projektu

Przedmiot: Warsztaty z technik uczenia maszynowego

Maciej Chrabąszcz
Weronika Kamińska
Aleksander Kozłowski
Michał Rosiński

Cel projektu

Porównanie klasyfikacji różnych rodzajów ryżu na dwa sposoby - jedno z podejść będzie opierało się na klasycznych algorytmach uczenia maszynowego, drugie zaś na sieciach neuronowych.

Dane

1. Źródła danych:

- a) <https://www.kaggle.com/mkoklu42/rice-image-dataset> – obrazy
- b) <https://www.kaggle.com/mkoklu42/rice-msc-dataset> – cechy

2. Opis danych:

Obydwa zbiory danych zawierają po 15000 przykładów dla każdego z 5 gatunków ryżu: *Basmati*, *Arborio*, *Jasmine*, *Ipsala*, *Karacadag* – łącznie 75000 przykładów. Zbiór z cechami jest zdefiniowany przez 106 zmiennych objaśniających takich jak np. powierzchnia, obwód, kurtoza. W drugim zbiorze danych znajdują się obrazy pojedynczych ziarenek ryżu. Zdjęcia są w formacie RGB i mają rozdzielczość 250x250 (poniżej przykładowe obrazki).



Kwestie techniczne

1. Język programowania: Python,
2. Wykorzystywane biblioteki: Numpy, Pandas, Pytorch, Scikit-Learn.

Wykorzystywane metody uczenia maszynowego

1. Klasyfikacja na podstawie cech:

Przeanalizowane zostaną zmienne znajdujące się w zbiorze danych. W ramach tworzenia rozwiązania określimy, które zmienne są istotne. Przetrenowane zostaną różne modele do klasyfikacji (między innymi KNN, model liniowy, drzewo decyzyjne, lasy losowe, naiwny Bayes). Przetestowane zostanie PCA i sprawdzone zostanie, ile składowych głównych wystarczy do klasyfikacji na wysokim poziomie.

2. Podejście Computer Vision:

Planujemy wykorzystać neuronowe sieci konwolucyjne. W celu testowania modeli dokonamy krosvalidacji. Planujemy również sprawdzić, czy moglibyśmy zmniejszyć zbiór danych uczących, wykorzystując augmentacje danych np. przesunięcie, wycinanie, obrót.

Podział zadań

- Klasyfikacja na podstawie zdjęcia – Maciej Chrabąszcz i Aleksander Kozłowski
- Klasyfikacja na podstawie cech – Weronika Kamińska i Michał Rosiński

