|  |  |
| --- | --- |
| Zbigniew Nowacki 234102  Bartosz Jurczewski 234067 | Rok akademicki 2019/20  Poniedziałek, 12:30 |

JĘZYKI PROGRAMOWANIA W ANALIZIE  
DANYCH – LABORATORIUM

Zadanie 3

# Opis implementacji

Zadanie zostało zrealizowane przy użyciu języka Python w wersji 3.7, z wykorzystaniem bibliotek: *pandas*, *matplotlib*, *numpy* oraz *sklearn*.

# Zbiór danych

Wybranym przez nas zbiorem zadań był zbiór „Leaf Classification”.

Zestaw danych zawiera około 1584 zdjęć okazów liści (16 próbek po 99 gatunków), które zostały przekonwertowane na binarne czarne liście na białym tle.

Zbiór zawiera trzy zestawy cech na obraz:

* ciągły deskryptor kształtu, (ang. *a shape contiguous descriptor*, kolumny **shape\_X**)
* histogram tekstury wewnętrznej (ang. *interior texture histogram*, kolumny **texture\_X**)
* histogram marginesu o małej skali (ang. *ﬁne-scale margin histogram*, kolumny **margin\_X**)

Gdzie x to liczba od 1 do 64.

Dla każdej cechy podano 64-atrybutowy wektor na próbkę liścia.

Zbiór dostępny jest pod następującym adresem: <https://www.kaggle.com/c/leaf-classification/data>

# Klasyfikacja