Bartłomiej Łoś, Martyna Mąka

**Klasyfikacja obrazu satelitarnego na podstawie klas pokrycia terenu**

Powstały projekt dotyczył klasyfikacji obrazu satelitarnego w odniesieniu do klas pokrycia terenu (wody, terenów zalesionych, zabudowy oraz użytków rolnych). Dane satelitarne zostały pobrane ze strony EarthExplorer, natomiast dane referencyjne pozyskane zostały z BDOT10k. Data wykonania zdjęć satelitarnych to 01.05.2023r.

Pierwsze 10 linijek kodu zostały załadowane biblioteki. Są to m.in. biblioteki do przetwarzania danych przestrzennych (osmdata, raster, sf, rgdal), do analizy danych (dplyr, mlr3) oraz do wizualizacji (mlr3viz).

Następnie wczytane zostały dane satelitarne w formacie TIFF dla kanałów RGB(red, green, blue). Dla każdego kanału tworzony jest obiekt typu „raster”, a następnie zostały one ze sobą połączone za pomocą funkcji ‘’stack” do jednego obiektu nazwanego ‘’rgb”. Służy to do tego, aby pracować z tymi danymi jako jednym obiektem. Do zdefiniowania nazwy każdego kanału użyta została funkcja ‘’names”.

Obraz zawierający sztuka

Opis wygenerowany automatycznie  
Ryc. Obraz satelitarny wyświetlony za pomocą funkcji ‘’plot’’.

Przygotowanie danych referencyjnych – wczytanie plików shapefile zawierających różne klasy gruntów (wody, tereny zalesione, użytki rolne, zabudowa), które będą używane jako dane referencyjne do klasyfikacji obrazu satelitarnego. Wczytane dane przetwarzamy następnie za pomocą funkcji st\_read z biblioteki sf, która zwraca dane w postaci obiektu typu sf , który zawiera geometrię obiektów przestrzennych oraz ich atrybuty.

W kolejnym kroku dane referencyjne zostały połączone za pomocą funkcji ‘’rbind” do jednego obiektu nazwanego ‘’reference\_data”. W celu sprawdzenia poprawności geometrii danych referencyjnych, używana jest funkcja "st\_is\_valid". Wynik tej funkcji wskazuje, czy dane referencyjne są poprawne.

W dalszej części kodu, dane referencyjne są uproszczone za pomocą funkcji "ms\_simplify" z pakietu "rmapshaper". Uproszczenie polega na zmniejszeniu liczby wierzchołków dla każdej geometrii, co przyspieszy późniejsze operacje na tych danych. Następnie, dane referencyjne są próbkowane za pomocą funkcji "st\_sample" z pakietu "sf". Liczba próbek jest określana przez argument "size". Funkcja służy do tego, aby uzyskać mniejszy zestaw danych do przeprowadzenia klasyfikacji obrazu satelitarnego. Funkcja ta pozwala wybrać liczbę próbek, którą użytkownik chce pobrać.

Następnie, funkcją "extract" z pakietu "raster", wyodrębniane są piksele odpowiadające próbkom danych referencyjnych. Wynik tej funkcji przechowywany jest jako obiekt typu "data.frame".

W kolejnym kroku, dane wyodrębnione są łączone za pomocą funkcji "cbind" z danymi referencyjnymi, a następnie za pomocą funkcji "st\_join" z oryginalnymi danymi referencyjnymi. Ostatnia operacja usuwa z danych referencyjnych kolumnę "osm\_id".

Następnie użyta funkcja ‘’class” wyświetla informacje o klasie obiektu, który znajduje się w kolumnie ‘’klasa’’. Funkcja ‘’factor” zmienia wartości w tej kolumnie na reprezentację w postaci czynnikowej, która służy do reprezentowania wartości jako kategorii.

W kolejnym kroku zostały dodanie kolumny x oraz y. Posłużyła do tego funkcja ‘’st\_coordinates’’ z pakietu ‘’sf’’, która służy do wyodrębniania współrzędnych geograficznych. Następnie obie kolumny dodane zostały do ramki danych ‘’refrence\_data5”, a kolumna ‘’geom’’ zostaje usunięta. Zmieniona zostaje również numeracja wierszy, a wynik zostaje przypisany do zmiennej ‘’reference\_data6”.

Po wykonaniu powyższych operacji na danych, zostaje utworzone zadanie uczenia maszynowego. Funkcja użyta, aby to wykonać to ‘’ mlr3spatiotempcv::TaskClassifST”. W tym wypadku ‘’reference\_data6” zostaje przypisane do atrybutu funkcji ‘’backend’’ i służy jako źródło danych. Klasy obiektów zostają ustawione jako ‘’target’’, czyli klasę obiektów, na której użytkownik chce przewidywać. Jako nazwy kolumn ze współrzędnymi geograficznymi zostają zdefiniowane kolumny ‘’x’’ i ‘’y’’. Również ‘’coords\_as\_features’’ zostaje ustawione na FALSE, co oznacza, że koordynaty nie będą brane pod uwagę jako cechy.

Następnie zostaje zdefiniowany model, a funkcją niezbędną do tej operacji była ‘’mlr3::lrn”. Algorytm klasyfikacji to ‘’classif.ranger’’, natomiast atrybut ‘’predict\_type’’ zostaje ustawiony na ‘’prob’’, co oznacza, że w tym miejscu model będzie przewidywać prawdopodobieństwa przynależności do poszczególnych klas.

W kolejnym kroku zostaje zdefiniowany oraz przeprowadzony resampling, gdzie funkcje, które były do tego niezbędne to ‘’mlr3::rsmp” i ‘’mlr3::resample”. ‘’Repeated\_spcv\_coords’’ oznacza powtarzanie ze specjalnym podziałem na podzbiory, które uwzględniają strukturę przestrzenną danych, a ‘’folds’’ i ‘’repeats’’ definiują liczbę podziałów i powtórzeń. W kolejnym kroku przeprowadzenia resamplingu, ‘’task’’ , ‘’learner’’ oraz utworzony obiekt ‘’resampling’’ zostają przekazane jako argumenty.

Kolejno obliczone zostają miary jakości modelu.

Wyniki:

Ocena modelu na podstawie walidacji krzyżowej jest równa 0.569. Oznacza to że wykorzystanie jedynie kanałów RGB jest niewystarczające do poprawnej klasyfikacji ukształtowania terenu.

Obraz zawierający tekst, diagram, Plan, mapa

Opis wygenerowany automatycznie