

# III Seminarium Python GeoData Science

Hel marzec 2020



Seminarium Python GeoData Science jest organizowane na Stacji Morskiej Instytutu Oceanografii UG w Helu przez Centrum GIS UG. Jest skierowane do wszystkich, którzy chcą nauczyć się wykorzystania sztucznej inteligencji do analizy danych przestrzennych przy pomocy otwartego oprogramowania. W trzech dwudniowych modułach poznamy spójny workflow bazujący na środowisku Jupyter Notebook, oprogramowaniu GIS (QGIS), języku programowania Python i platformie github. Poznamy praktycznie metody inżynierii danych przestrzennych oraz *machine and deep learning* wykorzystujące biblioteki Pythona.

**Termin: 8 – 13 marzec 2020**

Kurs: niedziela – piątek

9 – 11.00 Warsztaty

13.00 – 15.00 Warsztaty

17.00 – 19.00 Ćwiczenia (realizacja projektów samodzielnie lub w grupach)

Cena: 300 zł za moduł łącznie z noclegami, za **wszystkie trzy moduły 900 zł**, możliwość uczestniczenia tylko w wybranych modułach.

Zasadniczo brak jest wymagań wstępnych, ale przydatna będzie przeciętna znajomość Pythona lub QGISa.

## Moduł I - GeoPython w QGIS (8-9 marzec)

Pierwszego dnia będziemy realizować komplet zagadnień związanych z hasłem Geopython. Poznamy i nauczymy się pracować w Jupyter Notebook z QGISem. Będziemy ćwiczyć uruchamianie narzędzi QGIS z poziomu Pythona (processing) oraz czytanie i zapisywanie danych wektorowych, operacje na geometrii danych (biblioteka OGR), czytanie i zapisywanie danych rastrowych (biblioteka GDAL). Poznamy także metody algebry map oparte o NumPy i SciPy naukowe biblioteki Pythona. Poznamy wreszcie podstawowy pakiet wizualizacji danych (Matplotlib).

## Moduł II - Inżynieria danych przestrzennych i klasyfikacja nienadzorowana (10-11 marzec)

W tym module nauczymy się tworzenia i pracy z *tidy data* – podstawową strukturą danych w data

science. Podstawowym narzędziem pracy z danymi w środowisku Pythona jest Pandas. Zdobyć podstawowych umiejętności pracy z tą biblioteką będzie głównym celem pierwszego dnia. Zajmiemy się także metodami pozyskiwania danych, automatyzacją pobierania danych przez Internet oraz pozyskiwania danych z uwzględnieniem relacji przestrzennych. Uzupełnieniem inżynierii danych będzie eksploracyjna analiza danych (za pomocą pakietów Matplotlib i Seaborn). Następnie zajmiemy się analizą struktury danych pozwalającą na redukcję wymiarowości danych (atrybutów) za pomocą takich metod jak (PCA, t-SNE). Operacje te będziemy prowadzili w podstawowym pakiecie uczenia maszynowego (Scikit-learn). Ostatnim elementem tego modułu będzie klasyfikacja nienadzorowana (*K – means clustering, HCA*).

## Moduł III - Machine and deep learning. Klasyfikacja nadzorowana, sieci neuronowe (12-13 marzec)

W tym module cały czas będziemy doskonalić posługiwanie się Pandas i Scikit-Learn oraz bibliotekami wizualizacji. Zajmiemy się budową modeli klasyfikacji nadzorowanej i regresji z wykorzystaniem następujących metod: Gaussian Naive Bayes, k-Nearest Neighbors, Linear Regression, Logistic Regression, Support Vector Machines (SVMs), Decision Trees, Random Forest oraz Neural networks (*deep learning*). Zapoznamy się z rozwiązaniami tworzenia modeli przestrzennych (*geographically weighted methods*). Wykorzystane zostaną biblioteki: Scikit-learn, Tensor Flow, Keras.

Zapraszamy, Centrum GIS UG