

# Analiza serii czasowych w Pythonie z wykorzystaniem danych obserwacji Ziemi

Satelity umożliwiają więcej niż obserwację tego, co znajduje się na powierzchni Ziemi. Zbiory zdjęć, wykonywanych cyklicznie przy każdym przelocie satelity, układają się w szeregi czasowe. Obserwowana jest zmienność środowiska, jak i infrastruktury. Dobór zobrazowań, narzędzi do ich przetwarzania oraz modelu umożliwiającego analizę interesującego zjawiska to wszystko, co jest potrzebne ekspertowi Data Science do predykcji przyszłości. W artykule przedstawiono kroki potrzebne do zbudowania modelu prognostycznego dla zmiennej meteorologicznej nazywanej Land Surface Temperature (temperatura powierzchni obiektów na Ziemi) przy użyciu Pythona i jego bibliotek. Wiedza zawarta w tekście pozwoli czytelnikowi na wyprowadzenie własnych modeli dla serii czasowych charakteryzujących się sezonowością.

## WSTĘP: ROZDZIELCZOŚĆ ROZDZIELCZOŚCI NIERÓWNA

Zdjęcia satelitarne kojarzą się tylko ze statycznymi obrazami powierzchni Ziemi. Z tego względu istotną zmienną jest *rozdzielczość przestrzenna*, czyli pokrycie powierzchni na piksel, które warunkuje to, do jakich celów mogą zostać użyte obrazy satelitarne. Nie mniej ważnym parametrem jest *rozdzielczość czasowa* wykonywania poszczególnych zobrazowań. Rozdzielczość taką podaje się w jednostce czasu, na którą przypada wykonanie fotografii tego samego wycinka powierzchni Ziemi przez satelitę, czyli cykl orbitalny. Ponowne przeloty satelitów, których dane udostępniane są za darmo i mają dobrą rozdzielczość przestrzenną, takich jak *Landsat 8* albo *Sentinel 2*, liczone są w tygodniowych odstępach czasu. Dla niektórych zastosowań to zbyt rozległe interwały... Szczególnie istotne są tutaj sytuacje kryzysowe, zarządzanie wielkoskalową i dynamiczną infrastrukturą albo meteorologia. Nawet w rolnictwie takie przeloty mogą stanowić problem, ponieważ nie ma żadnej gwarancji, że w dzień, kiedy satelita znajduje się nad polami, chmury nie zasłonią powierzchni Ziemi i w efekcie nie uzyska się żadnych wyników.

Z pomocą przychodzą satelity, które mają znacznie lepszą rozdzielczość czasową, takie jak *Aqua* i *Terra* z instrumentem MODIS (z ang. *Moderate-Resolution Imaging Spectroradiometer*), które wykonują zdjęcia danego obszaru raz na dzień/dwa dni. Gdzie tkwi haczyk? Ceną częściej aktualizacji jest pogorszona rozdzielczość przestrzenna, nawet dwa rzędy wielkości gorsza niż ta oferowana przez *Landsat 8* albo *Sentinel 2*. Nie oznacza to, że dane tego typu są nieprzydatne. Zmienia się raczej skala rozpatrywania zjawisk. Dla satelitów o dobrej rozdzielczości przestrzennej możemy mówić o powierzchni budynku, w przeciwieństwie do wymienionego wcześniej instrumentu MODIS, który obejmuje obszary dużych dzielnic i miast<sup>1</sup>.

1. Czy istnieją satelity o bardzo dobrej rozdzielczości przestrzennej i czasowej? Tak, jednakże są to instrumenty komercyjne i dane pozyskiwane z tych instrumentów są płatne, dlatego oprócz parametrów technicznych satelitów na uwagę należy mieć kwestie biznesowe: nie oplaca się kupować zdjęć z każdego dnia albo dla dużych obszarów, więc ostatecznie i tak możemy skończyć z mniejszą ilością danych i produkt może nie spełnić pokładanych w nim oczekiwań. Z drugiej strony zarządzanie kryzysowe, działania wywiadowcze albo analiza dynamicznych zjawisk przestrzennych w miastach może wymagać użycia danych komercyjnych.

W artykule skoncentrowano się na niekomercyjnych danych o bardzo dobrej rozdzielczości czasowej i słabej rozdzielczości przestrzennej, dlatego:

- Prezentowany „use case” dotyczy będzie zjawisk wielkoskalowych, w tym przypadku zmian temperatury powierzchni w Warszawie na przestrzeni lat, od roku 2000 do roku 2019.
- Do analizy wykorzystany zostanie statystyczny model ARIMA oraz jego pochodna SARIMA.
- Przedstawiony będzie sposób walidacji algorytmów do analizy serii czasowych.

Wykorzystane dane pochodzą z instrumentu MODIS satelity *Terra*<sup>2</sup>. Poszczególne etapy ściągania danych opisano w dodatku 1 do artykułu, w którym znajduje się również odnośnik do przykładowego skryptu ułatwiającego pobieranie tych zobrazowań.

## I CZĘŚĆ 1: PRZYGOTOWANIE DANYCH

Każdy projekt Data Science rozpoczynamy od przygotowania danych. Ten proces jest trudniejszy w przypadku danych z instrumentu MODIS niż z prezentowanych we wcześniejszych artykułach danych *Landsat 8* albo udostępnianych przez Europejską Agencję Kosmiczną zdjęć z satelitów *Sentinel 2*. Sceny MODIS zapisywane są w nietypowym dla ilustracji formacie HDF (z ang. *Hierarchical Data Format*).

W artykule wykorzystywany jest produkt o nazwie MOD11B3<sup>3</sup>. Jest to średnia miesięczna temperatura powierzchni Ziemi. Jeden kafelek zajmuje 1200 na 1200 kilometrów, a jeden piksel ma 5600 metrów kwadratowych. Kafelek zawiera 19 warstw (!), z czego wykorzystane zostaną dwie: *Daytime Land Surface Temperature* oraz *Nighttime Land Surface Temperature*, obie w formacie uint16. Po przemnożeniu ich

2. <https://lpdaac.usgs.gov/products/mod11b3v006/>

3. Wan, Z., Hook, S., Hulley, G. (2015). MOD11B3 MODIS/Terra Land Surface Temperature/Emissivity Monthly L3 Global 6km SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. doi: 10.5067/MODIS/MOD11B3.006.