



**Kierunek studiów: Geoinformacja**

**Nr albumu: 461791**

**Filip Ratajszczak**

**Wykrywanie farm fotowoltaicznych na podstawie  
danych teledetekcyjnych**

*My title*

Praca inżynierska napisana  
w Instytucie Geoekologii i Geoinformacji  
pod kierunkiem  
dr. hab. Jakuba Nowosada

Poznań, 2024



# Streszczenie

## Abstrakt

Streszczenie powinno przedstawiać skrótowo główny problem pracy i jego rozwiązanie. Możliwa struktura streszczenia to: (1) 1-3 zdania wstępu do problemu (czym się zajmujemy, dlaczego jest to ważne, jakie są problemy/luki do wypełnienia), (2) 1 zdanie opisujące cel pracy, (3) 1-3 zdania przedstawiające użyte materiały (dane) i metody (techniki, narzędzia), (4) 1-3 zdania obrazujące główne wyniki pracy, (5) 1-2 zdania podsumowujące; możliwe jest też określenie dalszych kroków/planów.

Słowa kluczowe: (4-6 słów /zwrotów opisujących treść pracy, które nie wystąpiły w tytule)

## Abstract

The abstract must be consistent with the above text.

Keywords: (as stated before)

# Spis treści

<b>Streszczenie</b>	<b>3</b>
<b>1 Wprowadzenie</b>	<b>5</b>
<b>2 Przegląd literatury</b>	<b>7</b>
2.1 Podrozdział . . . . .	8
<b>3 Materiały</b>	<b>9</b>
3.1 Zdjęcia satelitarne . . . . .	9
3.2 Ortofotomapa i mozaiki zdjęć satelitarnych . . . . .	11
3.3 Próbkki treningowe i testowe . . . . .	12
<b>4 Metody</b>	<b>13</b>
4.1 Przygotowanie danych . . . . .	13
4.2 Próbkki treningowe i testowe . . . . .	13
4.3 Uczenie maszynowe . . . . .	13
4.4 Oprogramowanie . . . . .	15
<b>5 Wyniki</b>	<b>17</b>
<b>6 Podsumowanie</b>	<b>19</b>

# **Rozdział 1**

## **Wprowadzenie**

Wprowadzenie powinno mieć charakter opisu od ogółu do szczegółu (np. trzy-pięć paragrafów). Pierwszy paragraf powinien być najbardziej ogólny, a kolejne powinny przybliżać czytelnika do problemu. Przedostatni paragraf powinien określić jaki jest problem (są problemy), który praca ma rozwiązać i dlaczego jest to (są one) ważne.

Wprowadzenie powinno być zakończone stwierdzeniem celu pracy. Dodatkowo tutaj może znaleźć się również krótki opis co zostało zrealizowane w pracy.



## Rozdział 2

# Przegląd literatury

Ten rozdział zawiera wyjaśnienie kontekstu pracy.

Pisząc ten rozdział proszę pomyśleć o osobach, które zupełnie nie znają opisywanej tematyki. Należy tutaj krok po kroku wyjaśnić podstawowe koncepcje, istotność problemu, wyniki poprzednich podobnych badań, itd. Ten rozdział obejmuje tylko kwestie, które już zostały wykonane przez inne osoby - nowe wyniki mają swoje miejsce w rozdziale 5.

Każda kwestia opisana w tym rozdziale powinna być cytowana. Dodatkowo cytowania odbywa się poprzez uzupełnienie pliku `thesis.bib` zapisem w formacie BibTeX, a następnie dodanie nazwy referencji poprzedzonej znakiem `@`. Przykładowo, zacytowanie książki *Geocomputation with R* odbywa się poprzez `(lovelace_geocomputation_2019)`.

W przypadku, gdy cytowanie zostało poprawnie wpisane oraz istnieje w pliku `thesis.bib` to bibliografia powinna się automatycznie wygenerować na końcu pracy.

W przypadku, gdy praca dyplomowa opisuje konkretny obszar to można po tym rozdziale stworzyć kolejny rozdział opisujący "obszar badań".

Ten i kolejne rozdziały mogą mieć także podrozdziały. Tworzenie podrozdziałów polega na stworzeniu nowej linii rozpoczynającej się od znaków `##` a następnie tytułu podrozdziału. Dodatkowo w postaci `{#sec-}` można dodać skrót nazwy rozdziału/podrozdziału umożliwiając odnośnienie się do niego używając operatora `[-@sec]`.

## **2.1 Podrozdział**

Przykładowo, “te kwestie zostały opisane w podrozdziale 2.1”. Zwróć uwagę, że w ten sposób automatycznie tworzony jest odnośnik w pliku PDF.



## Rozdział 3

# Materiały

### 3.1 Zdjęcia satelitarne

#### 3.1.1 Sentinel-1

#### 3.1.2 Sentinel-2

Misja Sentinel-2 stanowi inicjatywę Komisji Europejskiej, która jest operacyjnie prowadzona przez Europejską Agencję Kosmiczną (ang. *European Space Agency*, ESA) w ramach programu Copernicus. Celem tej misji jest dostarczanie obrazów satelitarnych, obejmujących trzynaście zakresów spektralnych o różnych rozdzielczościach przestrzennych: 10, 20 lub 60 metrów, zależnie od rejestrowanego kanału. Rozdzielczość czasowa misji Sentinel-2 wynosi pięć dni nad równikiem i zwiększa się wraz ze wzrostem szerokości geograficznej, osiągając dwa dni na średnich szerokościach geograficznych (Hejmanowska et al., [2020](#)).

Dane pozyskiwane przez satelity Sentinel-2 są dostępne na różnych poziomach przetworzenia, lecz najczęściej używane przy tworzeniu map pokrycia terenu i użytkowania ziemi (ang. *Land Use/Land Cover*, LULC) są produkty 1C (współczynnik odbicia na poziomie górnej części atmosfery; ang. *Top-of-Atmospheric reflectance*, TOA) oraz 2A (współczynnik odbicia na powierzchni Ziemi; ang. *Bottom-of-Atmospheric reflectance*, BOA) (Phiri et al., [2020](#)).

**Tabela 3.1:** *Kanały spektralne satelitów Sentinel-2*

Kanał	Nazwa kanału	Centralna długość fali [nm]	Zakres spektralny [nm]	Rozdzielczość przestrzen- na [m]
B01	Coastal Aerosol	443	433–453	60
B02	Blue	490	458–523	10
B03	Green	560	543–578	10
B04	Red	665	650–680	10
B05	Vegetation RedEdge	705	698–713	20
B06	Vegetation RedEdge	740	733–748	20
B07	Vegetation RedEdge	783	773–793	20
B08	NIR	842	785–900	10
B8A	NIR	865	855–875	20
B09	Water Vapour	945	935–955	60
B10	Cirrus	1375	1360–1390	60
B11	SWIR	1610	1565–1655	20
B12	SWIR	2190	2100–2280	20

Produkty poziomu 1C to dane poddane korekcjom radiometrycznym i geometrycznym, prezentowane jako sceny o powierzchni 100 km<sup>2</sup> (100 x 100 km) w projekcji UTM/WGS84 (ESA, 2015). Skuteczne wykorzystanie tych danych w zastosowaniach związanych z terenami lądowymi wymaga precyzyjnej korekcji zdjęć satelitarnych pod kątem efektów atmosferycznych (Main-Knorn et al., 2017). Produkty poziomu 2A powstają poprzez zastosowanie dodatkowej korekcji atmosferycznej dla danych poziomu 1C za pomocą procesora korekcji atmosferycznej Sen2Cor (Main-Knorn et al., 2017).

Z dostępnych kanałów spektralnych (Tabela 3.1) wykorzystano 10 zakresów, ponieważ pasma rejestrowane w rozdzielczości 60 metrów są przeznaczone głównie do korekcji atmosferycznych i detekcji chmur. Kanał 1 (443 nm) służy do korekcji wpływu aerozoli, kanał 9 (940 nm) do korekcji wpływu pary wodnej, a kanał 10 (1375 nm) do wykrywania chmur typu cirrus (Drusch et al., 2012).

### 3.1.3 Indeksy spektralne - ???

### 3.1.4 Tekstury obrazu - ???

## 3.2 Ortofotomapa i mozaiki zdjęć satelitarnych

Do lokalizacji oraz wektoryzacji istniejących farm fotowoltaicznych wykorzystano ortofotomapę udostępnianą przez Główny Urząd Geodezji i Kartografii oraz mozaiki zdjęć satelitarnych dostarczane przez podmioty komercyjne. W teledetekcji jednym z zastosowań mozaik obrazów satelitarnych jest tworzenie zestawów danych referencyjnych poprzez interpretację wizualną, np. w celu walidacji wyników klasyfikacji produktów pokrycia terenu (Lesiv et al., 2018). Do stworzenia zbioru danych testowych i treningowych wykorzystano ortomozaiki Google Satellite, Bing Aerial, Mapbox Satellite oraz Planet Basemaps, udostępniane w formie usług sieciowych (WMS, WMTS, XYZ Tiles). Ortomozaiki te są tworzone na podstawie komercyjnych zdjęć satelitarnych wykonywanych przez podmioty takie jak Maxar Technologies, Airbus czy Planet Labs. Ortofotomapa udostępniana przez GUGiK charakteryzuje się rozdzielczością przestrzenną 25 cm, a rozdzielczość przestrzenna wykorzystanych mozaik obrazów satelitarnych (poza Planet Basemaps) jest wyższa niż 1 m, np. dla mozaiki Bing Aerial udostępnianej przez firmę Microsoft wynosi ona 30-60 cm. Często jednak nie jest możliwe ustalenie dat wykonania zdjęć satelitarnych, które posłużyły do stworzenia konkretnej mozaiki zobrazowań satelitarnych (Lesiv et al., 2018). Mozaika tworzona przez Planet na podstawie zdjęć satelitarnych wykonywanych przez konstelację satelitów PlanetScope charakteryzuje się rozdzielczością przestrzenną 4,77 m na równiku, jednak w porównaniu do pozostałych wymienionych produktów jest tworzona z miesięczną oraz kwartalną częstotliwością. Pozwala to, pomimo niższej rozdzielczości przestrzennej na stworzenie zbioru danych testowych i treningowych na konkretny okres czasu ([planet-basemaps-product-specifications.pdf](#)). Mozaiki Planet Basemaps są tworzone na podstawie zdjęć wybieranych przy użyciu algorytmu, który wybiera najlepsze obrazy z katalogu Planet w określonym przedziale czasowym. Wybierając najlepsze obrazy, Planet jest w stanie tworzyć wysokorodzielcze mozaiki, które są dokładne radiometrycznie i przestrzennie, a także charakteryzują się zminimalizowanym wpływem czynników atmosferycznych ([planet-basemaps-product-specifications.pdf](#)).

?Rycina przedstawiająca jedną farmę na ortofotomapie, mozaikach satelitarnych i scenie Sentinel-2 (kompozycja RGB)

### **3.3 Próbki treningowe i testowe**

## Rozdział 4

# Metody

### 4.1 Przygotowanie danych

Resampling kanałów w rozdzielczości 20 m do 10 m (metoda=bilinear). Złączenie kanałów w wielokanałowy raster. Opisać pozostałe czynności wykonane z tymi danymi.

?rycina - schemat przygotowania danych ?rycina - workflow Sentinel-1

### 4.2 Próbkowanie treningowe i testowe

### 4.3 Uczenie maszynowe

Klasyfikacja obrazów w teledetekcji polega na grupowaniu komórek w niewielkie zestawy klas, aby komórki w tych samych klasach miały podobne właściwości (Ismail et al., 2009). Istnieje wiele różnych metod klasyfikacji danych teledetekcyjnych. Stosunkowo nowymi podejściami wykorzystywanymi w tym kontekście są metody oparte na sztucznej inteligencji, takie jak uczenie maszynowe (ang. *Machine Learning*, ML) lub uczenie głębokie (ang. *Deep Learning*, DL) (Hejmanowska et al., 2020).

Uczenie maszynowe stanowi obszar sztucznej inteligencji, koncentrujący się na opracowywaniu algorytmów i modeli statystycznych zapewniających systemom komputerowym możliwość automatycznego uczenia się z danych i wykonywania określonych zadań bez konieczności bezpośredniego programowania. W przypadku skomplikowanych i

złożonych zestawów danych nie jesteśmy w stanie odpowiednio ich zinterpretować oraz wydobyć poprawnych informacji po wizualnym przejrzeniu danych (Mahesh, 2019). Uczenie maszynowe jest wykorzystywane do uczenia maszyn efektywnego przetwarzania danych (Sindayigaya et al., 2022). Algorytmy uczenia maszynowego można podzielić na cztery główne podejścia: uczenie nienadzorowane (ang. *unsupervised learning*), uczenie nadzorowane (ang. *supervised learning*), uczenie częściowo nadzorowane (ang. *semi-supervised learning*) oraz uczenie przez wzmacnianie (uczenie posiłkowane, ang. *reinforcement learning*) (Sarker, 2021).

W ciągu ostatnich dwudziestu lat zaproponowano kilka różnych algorytmów uczenia maszynowego do klasyfikacji obrazów satelitarnych (Sheykhmousa et al., 2020), zazwyczaj wykorzystujące techniki klasyfikacji bez nadzoru i klasyfikacji nadzorowanej (Ismail et al., 2009).

Uczenie nienadzorowane analizuje nieoznakowane zbiory danych bez konieczności ingerencji człowieka. Uczenie bez nadzoru jest powszechnie stosowane do eksploracji danych, ekstrakcji cech generatywnych, identyfikacji istotnych trendów i struktur oraz grupowania wyników. Ta technika uczenia maszynowego jest najczęściej używana do grupowania (klastowania), redukcji wielowymiarowości (redukcji cech) oraz identyfikacji skojarzeń i relacji (Sarker, 2021).

Nadzorowane algorytmy uczenia maszynowego wykorzystują oznaczone dane treningowe do znajdowania powiązań pomiędzy różnymi zmiennymi. Proces uczenia nadzorowanego zachodzi, gdy określone cele mają zostać osiągnięte na podstawie konkretnego zestawu danych wejściowych (treningowych). Dwa główne typy uczenia nadzorowanego to klasyfikacja, która separuje dane, oraz regresja, która dopasowuje dane (Sarker, 2021).

W badaniu do klasyfikacji wykorzystano nadzorowaną metodę lasów losowych (ang. *Random Forest*, RF) (Breiman, 2001).

### 4.3.1 Metoda lasów losowych

Random Forest stał się jednym z najpopularniejszych klasyfikatorów uczenia maszynowego wykorzystywanych przez społeczność teledetekcyjną ze względu na dokładność jego klasyfikacji oraz wysoką wydajność (Belgiu et al., 2016; Sheykhmousa et al., 2020). Metoda lasów losowych nie jest wrażliwa na szumy ani przetrenowanie, ponieważ nie opiera się na ważeniu (Gislason et al., 2006). Dodatkowo lasy losowe są znacznie lżejsze obliczeniowo niż metody oparte na wzmacnianiu (ang. *boosting*) (Gislason et al., 2006) oraz mogą przetwarzać różne typy danych, w tym zobrażenia satelitarne i dane numeryczne (Talukdar et al., 2020).

Random Forest (RF) to algorytm uczenia maszynowego, będący rozszerzeniem idei drzew decyzyjnych (Hejmanowska et al., 2020). Random Forest jest klasyfikatorem zespołowym (ang. *ensemble classifier*), co oznacza, że wykorzystuje dużą liczbę drzew decyzyjnych w celu przezwyciężenia słabości pojedynczego drzewa (Maxwell et al., 2018).

... do opisanie sposób działania (ciężkie)...

## 4.4 Oprogramowanie

### 4.4.1 QGIS

do wektoryzacji – tworzenia danych treningowych/testowych

### 4.4.2 SNAP

w przypadku wykorzystania danych Sentinel-1 – wykorzystany do przetwarzania danych Sentinel-1

### 4.4.3 Środowisko języka R

Czynności związane z końcowym przygotowaniem danych wejściowych oraz bezpośrednio z uczeniem maszynowym zostały wykonane z wykorzystaniem środowiska języka R (R Core Team, 2023). R to wieloplatformowy język programowania o otwartym kodzie

źródłowym do obliczeń statystycznych i wizualizacji danych (Lovelace et al., 2019). Dzięki dużej liczbie pakietów R obsługuje również statystyki geoprzestrzenne, modelowanie oraz wizualizację danych przestrzennych (Lovelace et al., 2019). W pracy wykorzystane zostało zintegrowane środowisko programistyczne (ang. *Integrated Development Environment*, IDE) RStudio (RStudio Team, 2020) przeznaczone dla języka R. Poza standardowymi możliwościami środowiska R, w procesie pracy wykorzystane zostały pakiety stworzone przez społeczność R w celu rozszerzenia funkcjonalności tego języka. Do operacji na danych rastrowych zastosowano pakiet *terra* (Hijmans, 2023), natomiast do przetwarzania danych wektorowych używany był pakiet *sf* (Pebesma, 2023). Losowe generowanie danych przestrzennych umożliwia pakiet *spatstat.random* (Baddeley et al., 2023b) z rodziny pakietów *spatstat* (Baddeley et al., 2023a). Do przeprowadzenia analizy oraz predykcji opartej o elementy uczenia maszynowego wykorzystano pakiet *mlr3* (Lang et al., 2023), w ramach którego użyty został algorytm lasów losowych zaimplementowany w pakiecie *ranger* (Wright et al., 2023). Do obliczeń związanych z uczeniem maszynowym wykorzystano pakiet *future* (Bengtsson, 2023), umożliwiający równoległe (wielowątkowe) przetwarzanie wyrażeń R, pozwalające na skrócenie czasu realizacji zadań w stosunku do przetwarzania sekwencyjnego.

?GLCMTextres jeśli dojdą tekstury obrazu



## Rozdział 5

# Wyniki

Część **Wyniki** może składać się z jednego lub więcej rozdziałów. Każdy z tych rozdziałów powinien mieć tytuł adekwatny do swojej treści.

Rozdziały wynikowe powinny korzystać z wiedzy opisanej w poprzednich rozdziałach (Rozdziały 2, 3, 4). W przypadku prac analitycznych, ich treść powinna przedstawiać kolejne etapy eksploracji i analizy danych. W przypadku prac technicznych, treść tych rozdziałów powinna opisywać stworzone narzędzia, a następnie pokazywać ich zastosowanie/a.

W przypadku prac technicznych warto pokazywać fragmenty napisanego rozwiązania lub jego wywołania używając bloków kodu.

```
moja_funkcja = function(x){  
  cat(x, "rządzi!")  
}  
moja_funkcja("Autor tej pracy")
```

Autor tej pracy rządzi!



## **Rozdział 6**

# **Podsumowanie**

Podsumowanie pracy jest w pewnym sensie znacznie rozbudowanym abstraktem. Należy wyliczyć i opisać osiągnięcia uzyskane w pracy dyplomowej. Tutaj jednak (w przeciwieństwie do np. rozdziału 1) należy przechodzić od szczegółu do ogółu - co zostało stworzone/określone, jak zostało to zrobione, jakie ma to konsekwencje, itd.

Ten rozdział powinien też zawierać opis kwestii, których nie udało się rozwiązać w pracy dyplomowej (i dlaczego się nie udało) oraz pomysły na przyszłe ulepszenie uzyskanych wyników lub dalsze badania.



# Bibliografia

- Baddeley, A, R Turner, and E Rubak (2023a). *spatstat: Spatial Point Pattern Analysis, Model-Fitting, Simulation, Tests*. R package version 3.0-7. <http://spatstat.org/>.
- Baddeley, A, R Turner, E Rubak, and T Davies (2023b). *spatstat.random: Random Generation Functionality for the spatstat Family*. R package version 3.2-1. <http://spatstat.org/>.
- Belgiu, M and L Drăguț (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing* **114**, 24–31.
- Bengtsson, H (2023). *future: Unified Parallel and Distributed Processing in R for Everyone*. R package version 1.33.0. <https://CRAN.R-project.org/package=future>.
- Breiman, L (2001). Random Forests. *Machine Learning* **45**, 5–32.
- Drusch, M, U Del Bello, S Carlier, O Colin, V Fernandez, F Gascon, B Hoersch, C Isola, P Laberinti, P Martimort, A Meygret, F Spoto, O Sy, F Marchese, and P Bargellini (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment* **120**. The Sentinel Missions - New Opportunities for Science, 25–36.
- ESA, ESA (2015). *Sentinel-2 User Handbook*. 64 pages. [https://sentinels.copernicus.eu/web/sentinel/user-guides/document-library/-/asset\\_publisher/xls1t4309D5h/content/sentinel-2-user-handbook](https://sentinels.copernicus.eu/web/sentinel/user-guides/document-library/-/asset_publisher/xls1t4309D5h/content/sentinel-2-user-handbook).
- Gislason, PO, JA Benediktsson, and JR Sveinsson (2006). Random Forests for land cover classification. *Pattern Recognition Letters* **27**(4). Pattern Recognition in Remote Sensing (PRRS 2004), 294–300.
- Hejmanowska, B and P Wezyk (2020). *Dane satelitarne dla administracji publicznej*. Polska Agencja Kosmiczna.

- Hijmans, RJ (2023). *terra: Spatial Data Analysis*. R package version 1.7-39. <https://rspatial.org/>.
- Ismail, MH, H Pakhriazad, and M Shahrin (2009). Evaluating supervised and unsupervised techniques for land cover mapping using remote sensing data. *Geografia : Malaysian Journal of Society and Space*.
- Lang, M, B Bischl, J Richter, P Schratz, M Binder, F Pfisterer, R Sonabend, and M Becker (2023). *mlr3: Machine Learning in R - Next Generation*. R package version 0.17.0. <https://CRAN.R-project.org/package=mlr3>.
- Lesiv, M, L See, JC Laso Bayas, T Sturn, D Schepaschenko, M Karner, I Moorthy, I McCallum, and S Fritz (2018). Characterizing the Spatial and Temporal Availability of Very High Resolution Satellite Imagery in Google Earth and Microsoft Bing Maps as a Source of Reference Data. *Land* 7(4).
- Lovelace, R, J Nowosad, and J Muenchow (2019). *Geocomputation with R*. CRC Press.
- Mahesh, B (2019). Machine Learning Algorithms - A Review. en. 9(1).
- Main-Knorn, M, B Pflug, J Louis, V Debaecker, U Müller-Wilm, and F Gascon (2017). Sen2Cor for Sentinel-2. In: *Image and Signal Processing for Remote Sensing XXIII*. Ed. by L Bruzzone. Vol. 10427. International Society for Optics and Photonics. SPIE, pp.37–48. <https://doi.org/10.1117/12.2278218>.
- Maxwell, AE, TA Warner, and F Fang (2018). Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing* 39(9), 2784–2817. eprint: <https://doi.org/10.1080/01431161.2018.1433343>.
- Pebesma, E (2023). *sf: Simple Features for R*. R package version 1.0-13. <https://CRAN.R-project.org/package=sf>.
- Phiri, D, M Simwanda, S Salekin, VR Nyirenda, Y Murayama, and M Ranagalage (2020). Sentinel-2 Data for Land Cover/Use Mapping: A Review. *Remote Sensing* 12(14).
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>.
- RStudio Team (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC. Boston, MA. <http://www.rstudio.com/>.
- Sarker, I (2021). *Machine Learning: Algorithms, Real-World Applications and Research Directions*.

- Sheykhmousa, RM and M Mahdianpari (2020). Support Vector Machine vs. Random Forest for Remote Sensing Image Classification: A Meta-analysis and systematic review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
- Sindayigaya, L and A Dey (2022). Machine Learning Algorithms: A Review. *Information Systems Journal*, vol. ISJ-RA-3392, sp 6.
- Talukdar, S, P Singha, S Mahato, Shahfahad, S Pal, YA Liou, and A Rahman (2020). Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. *Remote Sensing* **12**(7).
- Wright, MN, S Wager, and P Probst (2023). *ranger: A Fast Implementation of Random Forests*. R package version 0.16.0. <https://CRAN.R-project.org/package=ranger>.