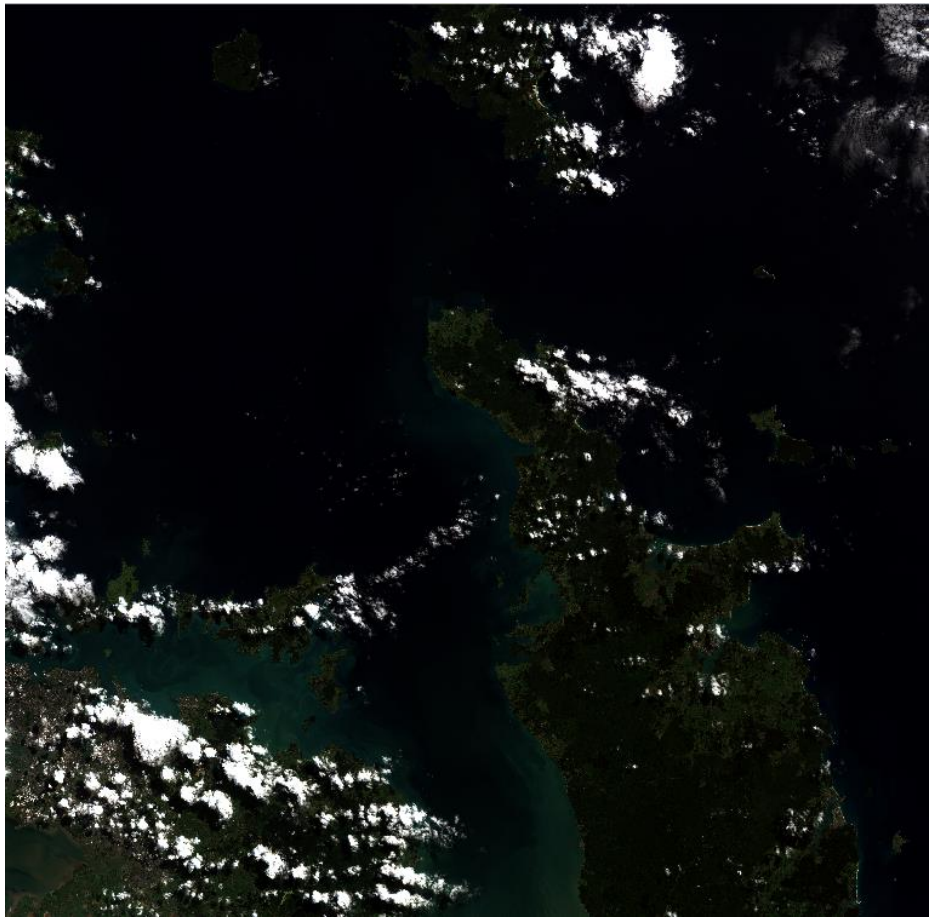
Klasyfikacja nadzorowana (Random Forest)

Kamil Gurlaga

Charakterystyka danych

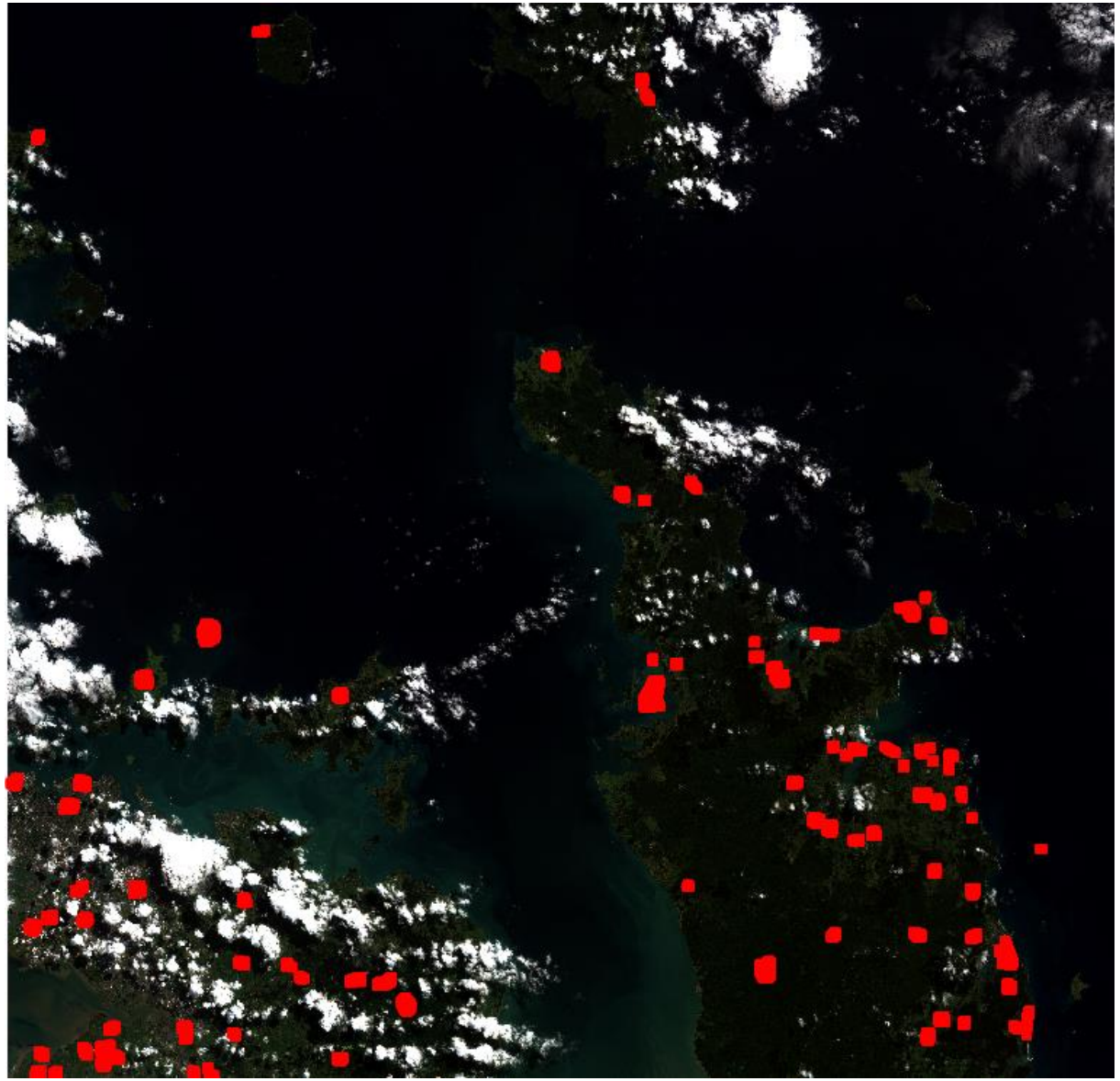


- Północ North Island, Nowa Zelandia
- Przewaga oceanu w pokryciu terenu
- Zdjęcia satelitarne rozdzielczość 20x20m

Zastosowane biblioteki

- import os
- import numpy as np
- import pandas as pd
- import seaborn as sns
- import matplotlib.pyplot as plt
- import rasterio as rio
- from sklearn.svm import SVC as SVM
- from sklearn.model_selection import train_test_split
- from sklearn.discriminant_analysis import LinearDiscriminantAnalysis as LDA
- from sklearn.ensemble import RandomForestClassifier as RFC
- from sklearn.metrics import confusion_matrix
- from sklearn.metrics import roc_curve, accuracy_score, f1_score, recall_score, precision_score, auc
- from sklearn.metrics import precision_recall_curve
- from sklearn.model_selection import learning_curve
- from sklearn.model_selection import ShuffleSplit
- from sklearn.model_selection import cross_val_score
- import warnings
- from sklearn.preprocessing import StandardScaler
- from sklearn.model_selection import GridSearchCV
- import os
- from rasterio.transform import from_origin

Próbki



Macierze

Zbiór treningowy

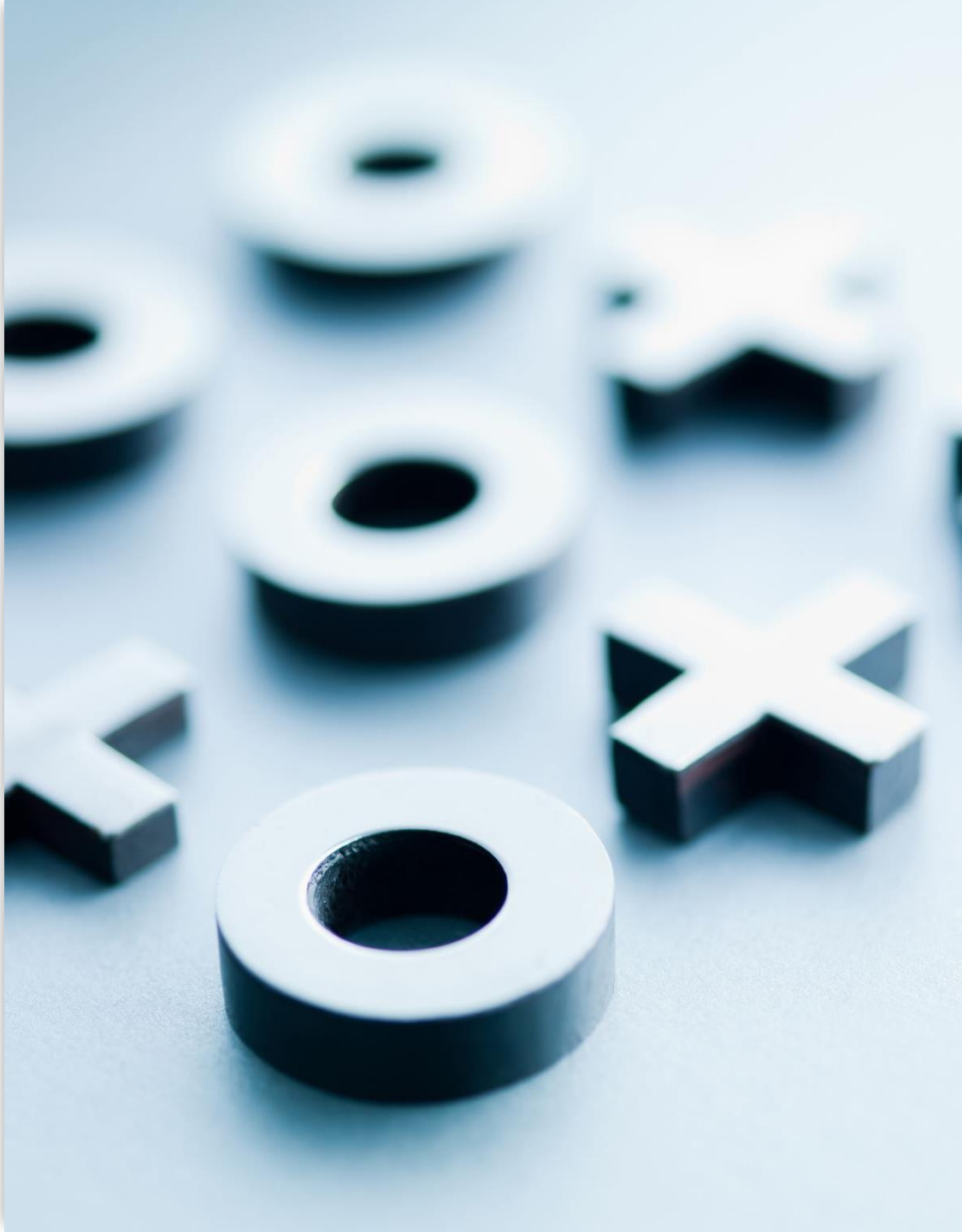
```
array([[ 7082,    56,     0,   265,    42],
       [   20,  5513,     0,   216,  1882],
       [    0,     0,  7518,     0,     2],
       [  372,    31,     0, 11157,   116],
       [   90,  1569,     0,   595,  5070]], dtype=int64)
```

Zbiór testowy

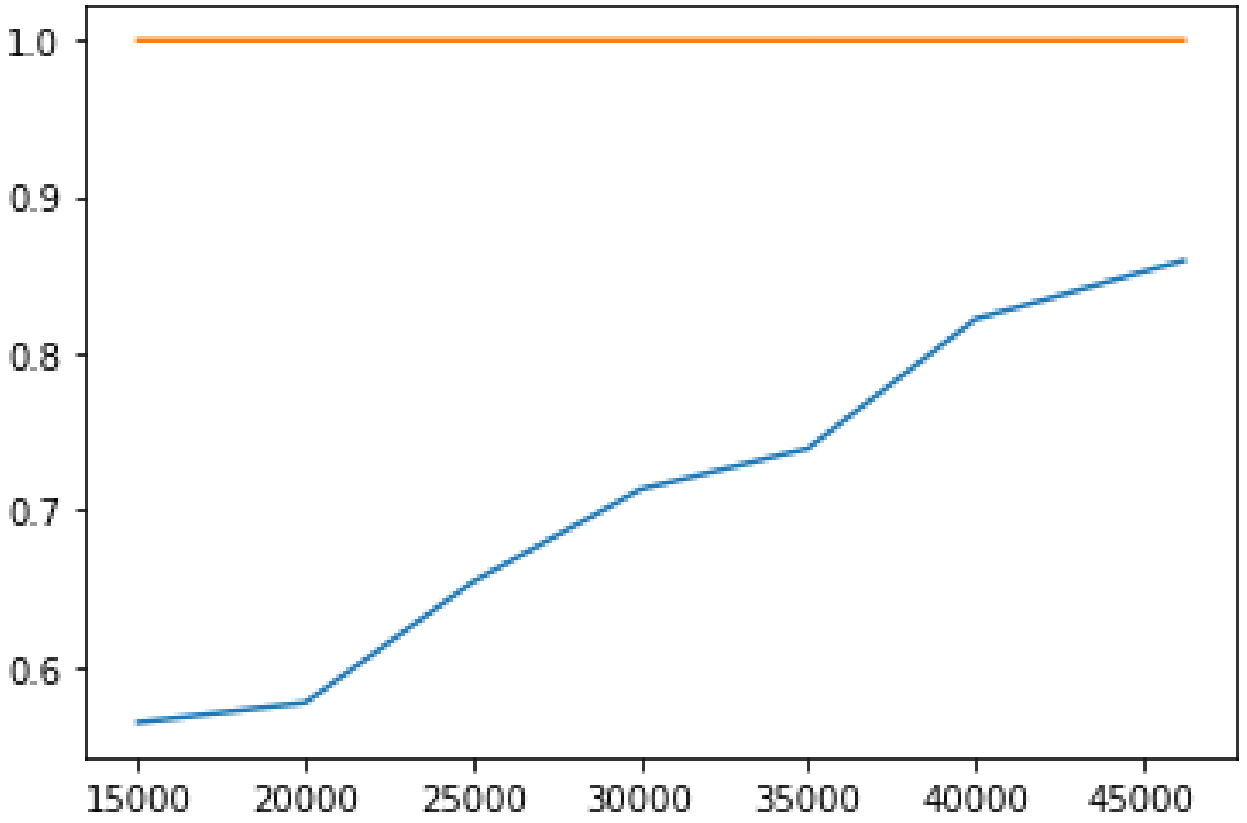
```
array([[4720,   49,     0,  243,   47],
       [  10, 3528,     0,  144, 1236],
       [   0,    0,  4991,    0,    1],
       [ 277,   29,     0, 7527,   82],
       [  82,  996,     3,  386, 3381]], dtype=int64)
```

scores

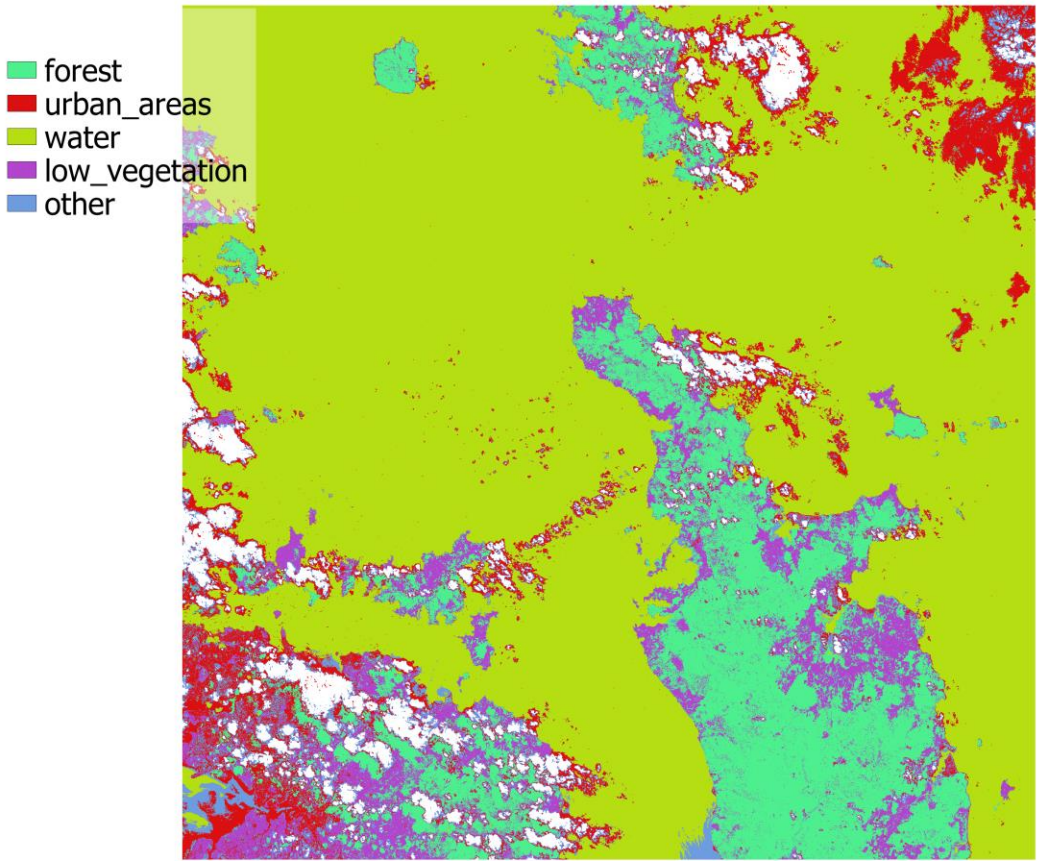
- Wartości wskaźnika cross_val_scores wyniosły: **[0.82662628 0.87047454 0.88374441 0.94915254 0.88012982]**
- Najlepsze parametry dla modelu to **{'max_depth': 5, 'min_samples_split': 2, 'n_estimators': 40}**



Wielkość zbioru testowego



Wynik



- Maska chmur niedokładna
- Lepszy dobór próbek
- Większa kontrola nad błędami I i II typu