

# Analiza statystyczna przyznawania funduszy UE gminom

Krzysztof Rudnicki, Michał Śar

June 5, 2024

## Contents

1	Wstęp	1
2	Omówienie rozdziałów	2
3	Opis literatury	2
4	Proces badawczy	3

## Abstract

This is a brief summary of your study, its results, and major conclusions.

## 1 Wstęp

**Kontekst** W 2024 mija 20 lat od wstąpienia Polski do Unii Europejskiej ?. Od tamtej pory bilans Polski w stosunku do Brukseli wynosi 175 miliardów euro na plus dla Polski ? W samym 2023 roku Polska otrzymała z UE prawie 3.5 miliarda złotych, wpłacając niecały miliard złotych ? W naszej pracy ponawiamy analizę statystyczną wykonaną sprzed 7 lat, na nowych danych, od początku roku 2014 do końca roku 2023

**Cel** Celem pracy jest sprawdzenie, jakie dane na temat gminy najbardziej korelują z liczbą przyznanych funduszy Unii Europejskiej danej gminie

**Hipoteza** Gęstość zaludnienia jest **najważniejszym** czynnikiem wpływającym na przyznanie rodków unijnych

### **Metoda badawcza**

1. Zebra dane UE
2. Zebra dane gmin
3. Poczy dane po numerze TERYT
4. Przeanalizowa dane
5. Wyświetli wyniki

### **Wyniki**

## **2 Omówienie rozdziałów**

Na początku artykułu przedstawiamy czemu wybraliśmy taki temat, co chcemy osiągnąć naszą pracę, w jaki sposób chcemy to osiągnąć i jaki rezultat ostatecznie udało nam się pokazać

Następnie opisujemy istniejącą literaturę na temat rodków Unijnych z którą się zapoznaliśmy i przedstawiamy w czym różni się nasza praca od istniejących

Potem tłumaczymy nasz proces badawczy, w jaki sposób zbieraliśmy i czytaliśmy dane, jak je analizowaliśmy i jak przedstawialiśmy wyniki

Kontynuując, pokazujemy co otrzymaliśmy ostatecznie w wyniku naszej pracy. Przedostatni rozdział zajmuje się dyskusją wyników, przedstawiamy co udało nam się osiągnąć i dlaczego, czego nie udało nam się osiągnąć i dlaczego oraz przede wszystkim konfrontujemy wynik z naszą hipotezą

Na końcu podsumowujemy całą pracę i przedstawiamy spis literatury z której korzystaliśmy

## **3 Opis literatury**

**Decision trees: from efficient prediction to responsible AI** Artykuł poświęcony jest omówieniu drzew decyzyjnych, rozpoczyna od zdefiniowania

czym drzewo decyzyjne jest, jakie s jego unikalne cechy, gdzie jest stosowane, jakie ma wady i potencjalne zagroenia oraz jak mona je zminimalizowa ? Wybraliśmy ten artykuł gdy opisuje jedn z głównych metod którą zamierzamy stosowa w naszym procesie badawczym do przeanalizowania danych

**Application of Successful EU Funds Absorption Models to Sustainable Regional Development** Artykuł wykorzystuje ankiety pytając 244 osób o to jak efektywnie wykorzystywane były fundusze UE w Polsce, Słowenii, Węgry i Chorwacji. Artykuł podkreśla znaczenie możliwości technicznych, administracyjnych, koordynacji pomiędzy instytucjami i dobrymi mechanizmami nadzorowania funduszy europejskich jako kluczowe dla skutecznego wykorzystywania funduszy unijnych. ?

Artykuł przyda się nam w ocenie jakie parametry pozytywnie wpływają na korzystanie z funduszy UE i jakie moglibyśmy mieć w naszym modelu. W naszym artykule zamiast ankiet wykorzystujemy dostępne już dane, a wyniki staramy się stworzyć przy użyciu modeli statystycznych. Dodatkowo zajmujemy się przedstawieniem jakie parametry wpływają na przyznanie środków UE a nie na to w jaki sposób można te środki skutecznie wykorzystywać

**Its not about the money. EU funds, local opportunities, and Euroscepticism** ) Artykuł opisuje jak pieniądze z Unii Europejskiej wpływają na eurosceptycyzm w danym kraju na podstawie Walii w kontekście referendum "Brexit". Badanie wykorzystuje metodę Regression discontinuity design (RDD), wybrano Wali z uwagi na różnicę w ilości pieniędzy przekazanych poszczególnym regionom. Autorzy wykazali że sama ilość pieniędzy przekazana danemu regionowi nie zwiększa znacznie poparcia dla Unii Europejskiej, natomiast duże nakłady powiązane z widoczną, namacalną poprawą na lokalnym rynku wpływają pozytywnie na postrzeganie Unii Europejskiej w lokalnych społecznościach ?

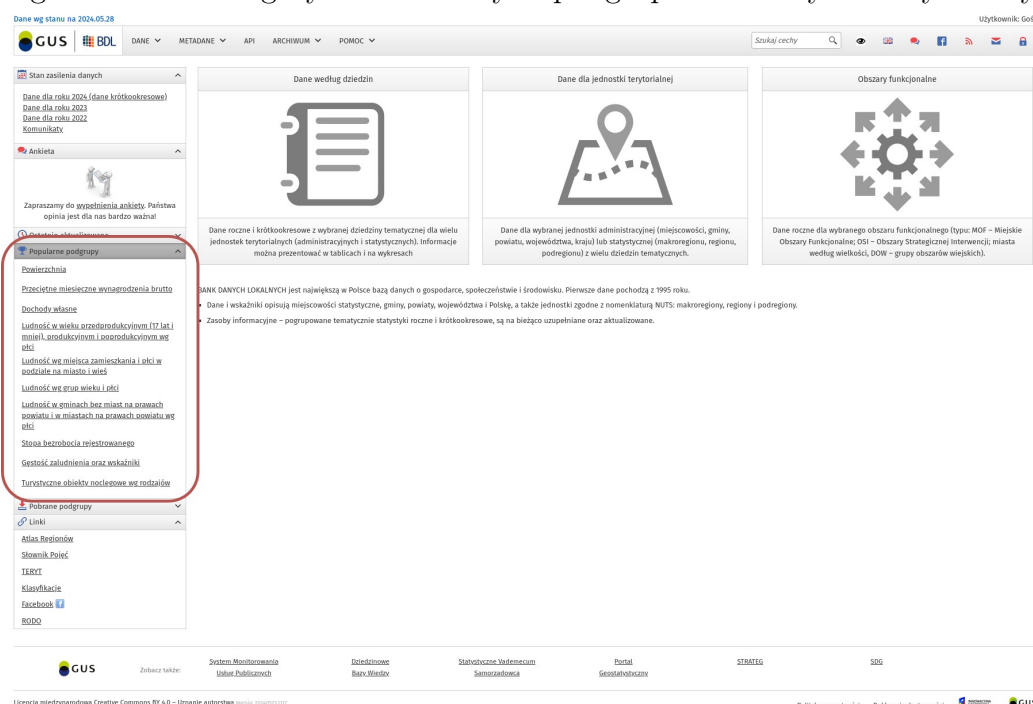
Nasz artykuł koncentruje się na tym co wpływa na przyznanie funduszy unijnych a nie na samą reakcję na ich przyznanie

## 4 Proces badawczy

Proces badawczy podzieliliśmy na 3 zasadnicze etapy, zebranie danych, przeanalizowanie ich i zaprezentowanie wyników

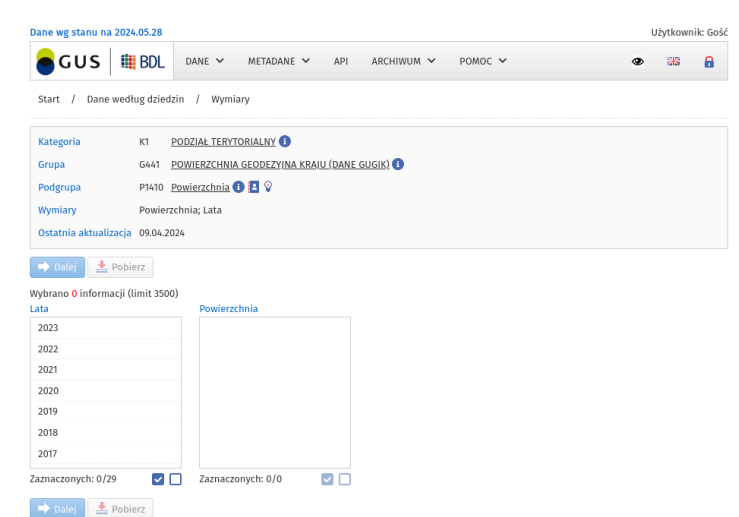
**Zbieranie danych** Wszystkie dane pobieraliśmy ze strony GUS-u <https://bdl.stat.gov.pl/bdl/start>  
Dane wybieraliśmy z zakładki "Popularne podgrupy", następnie wybieraliśmy wszystkie lata które nas interesowały (od 2014 do 2023 roku włącznie), po przejściu dalej wybieraliśmy wszystkie gminy, finalnie otrzymując tablicę którą pobieraliśmy do formatu csv

Figure 1: Strona z gusy z zaznaczonymi podgrupami z których korzystaliśmy



**Przygotowywanie danych** Dane musieliśmy przetworzyć przed ich wykorzystaniem, przede wszystkim usuwaliśmy wszelkie wiersze zawierające wartości puste lub NULL

Figure 2: Strona z gusy z zaznaczonymi podgrupami z których korzystaliśmy



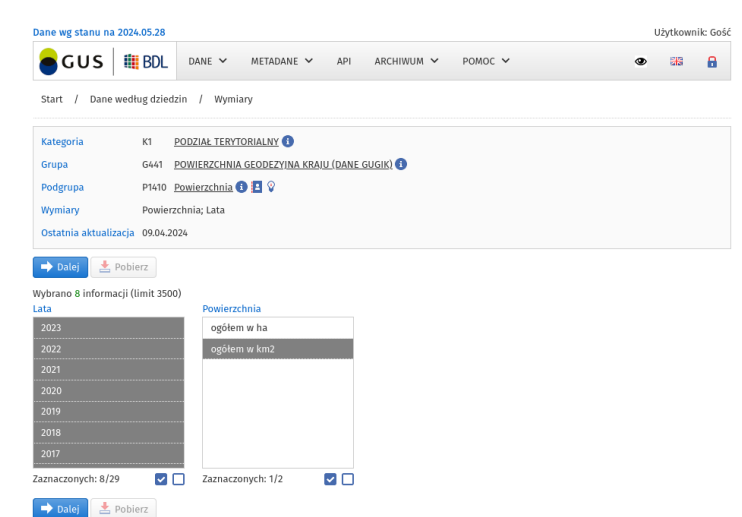
## Description of Code

The provided Python code snippet is part of a Jupyter Notebook. Its main functionality is to calculate and display the feature importance of a trained machine learning model. Feature importance is a key metric in understanding which features (or input variables) have the most significant impact on the model's predictions. Here's a breakdown of what the code does:

1. It creates a dictionary, `feature_importance`, which maps feature names to their respective importance scores as determined by the model.
2. The dictionary is then sorted by importance scores in descending order.
3. Finally, the code iterates through the sorted dictionary and prints out each feature's importance score alongside the feature's name.

The code snippet provided is as follows:

Figure 3: Strona z gusy z zaznaczonymi podgrupami z których korzystaliśmy



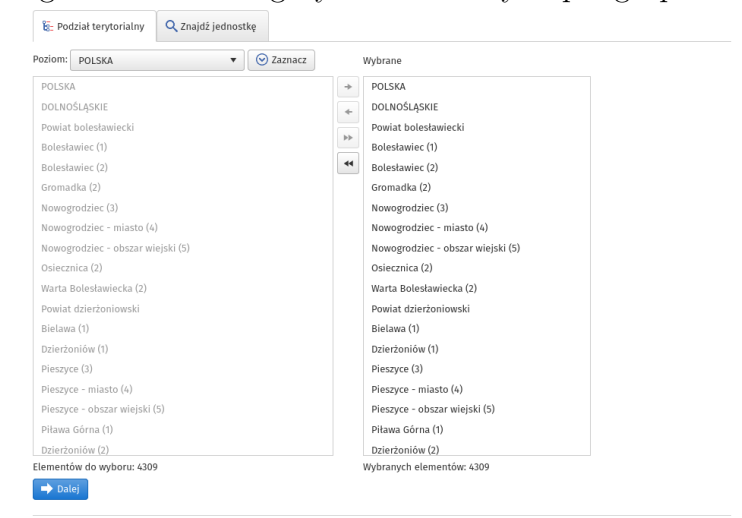
```
feature_importance = dict(zip(feature_names, model.
    feature_importances_))
for feature, importance in sorted(feature_importance.items(),
    key=lambda x: x[1], reverse=True):
    print(f'{importance:.5f} {feature}')
```

## Data Processed

The code processes the following types of data:

- `feature_names`: A list of feature names used in the model.
- `model.feature_importances_`: An array of feature importance scores generated by the model.

Figure 4: Strona z gusy z zaznaczonymi podgrupami z których korzystaliśmy



The purpose of this data processing is to identify and rank the most influential features in the model, which can be useful for feature selection, model interpretation, and improving model performance.

Figure 5: Strona z gusy z zaznaczonymi podgrupami z których korzystaliśmy

The screenshot shows a web application interface with a table of territorial units. The table has three columns: 'Jednostka terytorialna', '2017', and '2018'. The 'Export' menu is open, showing options for downloading the data in various formats.

Jednostka terytorialna ▲	2017	2018
POLSKA	312 679	
DOLNOŚLĄSKIE	19 947	19 947
Powiat bolesławiecki	1 304	1 304
Bolesławiec (1)	24	24
Bolesławiec (2)	288	288
Gromadka (2)	268	268
Nowogrodziec (3)	176	176
Nowogrodziec - miasto (4)	16	16
Nowogrodziec - obszar wiejski (5)	160	160
Osiecznica (2)	438	438
Warta Bolesławiecka (2)	110	110
Powiat dzierzoniowski	479	479
Bielawa (1)	36	36

Export menu options:

- XLS - tablica wielowymiarowa
- XLS - tablica relacyjna (zip)
- XLS - tablica przestawna
- CSV - tablica wielowymiarowa
- CSV - tablica relacyjna (zip)