

# Analiza statystyczna przyznawania funduszy UE gminom

Krzysztof Rudnicki, Michał Śar

6 czerwca 2024

## Spis treści

1	Wstęp	1
2	Omówienie rozdziałów	2
3	Opis literatury	3
4	Proces badawczy	4
5	Wyniki	9
6	Dyskusja	12
7	Konkluzja	13

## Streszczenie

Artykuł skupia się na stworzeniu modelu opisującego jakie parametry gminy najbardziej wpływają na przyznanie jej funduszy UE w latach 2014-2023.

## 1 Wstęp

**Kontekst** W 2024 roku mija 20 lat od wstąpienia Polski do Unii Europejskiej [5]. Od tamtej pory bilans Polski w stosunku do Brukseli wynosi 175 miliardów euro na plus dla Polski [4]. W samym 2023 roku Polska otrzymała

z UE prawie 3.5 miliarda złotych, wpacając niecały miliard złotych [3]. 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 środków unijnych.

### **Metoda badawcza**

1. Zebrać dane UE.
2. Zebrać dane gmin.
3. Pobrać dane po numerze TERYT.
4. Przeanalizować dane.
5. Wyświetlić wyniki.

## **2 Omówienie rozdziałów**

Na początku artykułu przedstawiamy, dlaczego 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 środków 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 są jego unikalne cechy, gdzie jest stosowane, jakie ma wady i potencjalne zagrożenia oraz jak można je zminimalizować [1]. Wybraliśmy ten artykuł, gdyż opisuje jeden z głównych metod, którą zamierzamy stosować 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 [6]. Artykuł przyda się nam w ocenie, jakie parametry pozytywnie wpływają na korzystanie z funduszy UE i jakie moglibyśmy znaleźć 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.

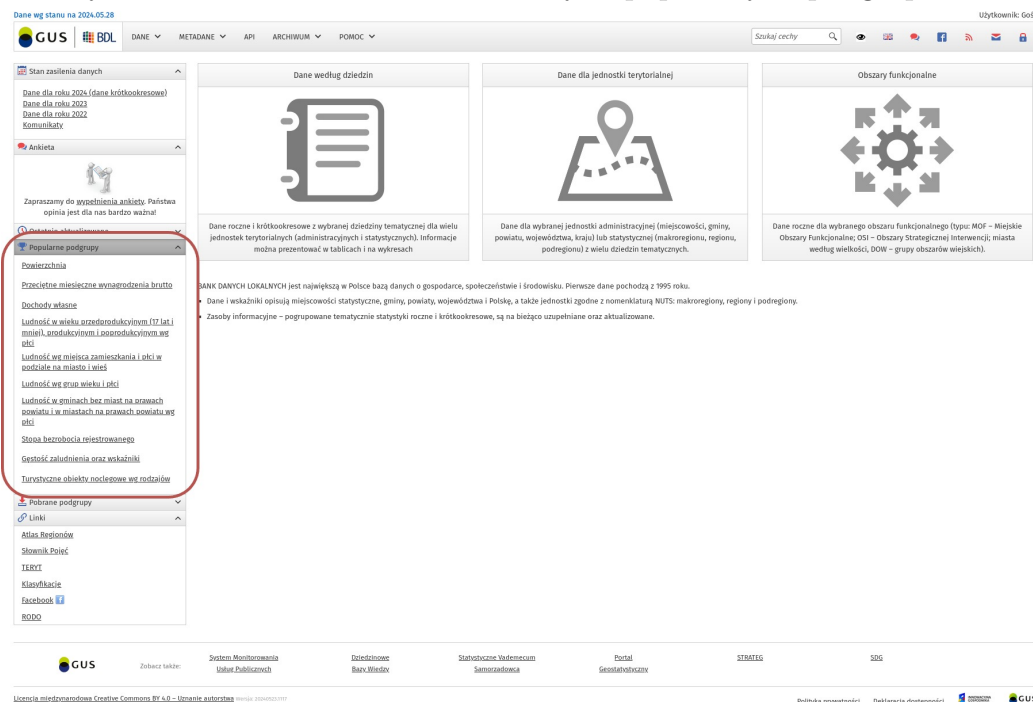
**It's not about the money. EU funds, local opportunities, and Euroscepticism** Artykuł opisuje, jak pieniądze z Unii Europejskiej wpływają na eurosceptycyzm w danym kraju na podstawie Walii w kontekście referendum "Brexit". Badanie wykorzystuje metodę Regression Discontinuity Design (RDD), wybrano Wali 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 widocznymi, namacalnymi poprawami na lokalnym rynku wpływają pozytywnie na postrzeganie Unii Europejskiej w lokalnych społecznościach [2]. 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 trzy zasadnicze etapy: zebranie danych, przeanalizowanie ich i zaprezentowanie wyników.

**Zbieranie danych** Wszystkie dane pobieraliśmy ze strony GUS-u: <https://bdl.stat.gov.pl/bdl/stat>. 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.



Rysunek 1: Strona GUS z zaznaczonymi popularnymi podgrupami



Rysunek 2: Strona GUS z zaznaczonymi latami i powierzchnią

Dane wg stanu na 2024.05.28

Użytkownik: Gość






DANE

METADANE

API

ARCHIWUM

POMOC



Start

/

Dane według dziedziny

/

Wymiary

Kategoria

K1

[PODZIAŁ TERYTORIALNY](#)

Grupa

G441

[POWIERZCHNIA GEODEZYJNA KRAJU \(DANE GUGIK\)](#)

Podgrupa

P1410

[Powierzchnia](#)

Wymiary

Powierzchnia; Lata

Ostatnia aktualizacja

09.04.2024

Dalej

Pobierz

Wybrano 8 informacji (limit 3500)

Lata

2023

2022

2021

2020

2019

2018

2017

Powierzchnia

ogółem w ha

ogółem w km2

Zaznaczonych: 8/29

☒

☐

Zaznaczonych: 1/2

☒

☐

Dalej

Pobierz

5

Rysunek 3: Strona GUS z zaznaczonymi powiatami

Podział terytorialny

Znajdź jednostkę

Poziom: POLSKA

Zaznacz

POLSKA

DOLNOŚLĄSKIE

Powiat bolesławiecki

Bolesławiec (1)

Bolesławiec (2)

Gromadka (2)

Nowogrodzic (3)

Nowogrodzic - miasto (4)

Nowogrodzic - obszar wiejski (5)

Osiecznica (2)

Warta Bolesławiecka (2)

Powiat dzierzoniowski

Bielawa (1)

Dzierżonów (1)

Pieszycy (3)

Pieszycy - miasto (4)

Pieszycy - obszar wiejski (5)

Piława Górna (1)

Dzierżonów (2)

→

←

→

←

Wybrane

POLSKA

DOLNOŚLĄSKIE

Powiat bolesławiecki

Bolesławiec (1)

Bolesławiec (2)

Gromadka (2)

Nowogrodzic (3)

Nowogrodzic - miasto (4)

Nowogrodzic - obszar wiejski (5)

Osiecznica (2)

Warta Bolesławiecka (2)

Powiat dzierzoniowski

Bielawa (1)

Dzierżonów (1)

Pieszycy (3)

Pieszycy - miasto (4)

Pieszycy - obszar wiejski (5)

Piława Górna (1)

Dzierżonów (2)

Elementów do wyboru: 4309

Wybranych elementów: 4309

➔ Dalej

Rysunek 4: Dane o powierzchni z moliwoci eksportu do CSV

The screenshot shows a web application interface with a table of territorial units. The 'Export' menu is open, showing options for XLS and CSV formats. The table displays data for various units, including Poland, Lower Silesia, and several counties and municipalities.

Jednostka terytorialna ▲	2017	2018
	[km <sup>2</sup> ]	[km <sup>2</sup> ]
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

**Przygotowywanie danych** Dane musimy przeprocesować przed ich wykorzystaniem, usuwaliśmy wiersze:

- Zawierające niepełny numer TERYT.
- Zawierające wartości Null albo puste.

Wybraliśmy w sumie 100 parametrów, na podstawie których ocenialiśmy wpływ na dotacje z UE. Można je podzielić na grupy:

1. Finansowe (dochody, wpływy, podatki).
2. Ludność (całkowita, na ppe, wiek przedprodukcyjny/produkcyjny/poprodukcyjny, gęstość zaludnienia).
3. Województwo.
4. Wymeldowania i zameldowania.

5. Turystyka.
6. Bezrobocie.
7. Typ gminy.
8. Odlego od Warszawy lub centrum decyzyjnego.

Dzielimy dane o dofinansowaniu UE na podstawie programów:

- Program Operacyjny Infrastruktura i środowisko 2014-2020.
- Program Operacyjny Inteligentny Rozwój.
- Program Operacyjny Polska Cyfrowa.
- Program Operacyjny Wiedza Edukacja Rozwój.
- Program Operacyjny Polska Wschodnia.

**Analiza danych** Wykorzystaliśmy model drzew decyzyjnych regresyjnych, wykorzystujących "Recursive Feature Elimination"(RFE). Trenowaliśmy model na głębokościach od 3 do 28 i na liczbie cech od 2 do 20. W ten sposób szukaliśmy najlepszego modelu, takiego, który wykazywa najmniejszy błąd MSE. Najlepsze parametry uzyskaliśmy dla głębokości 20 i liczby cech 13.

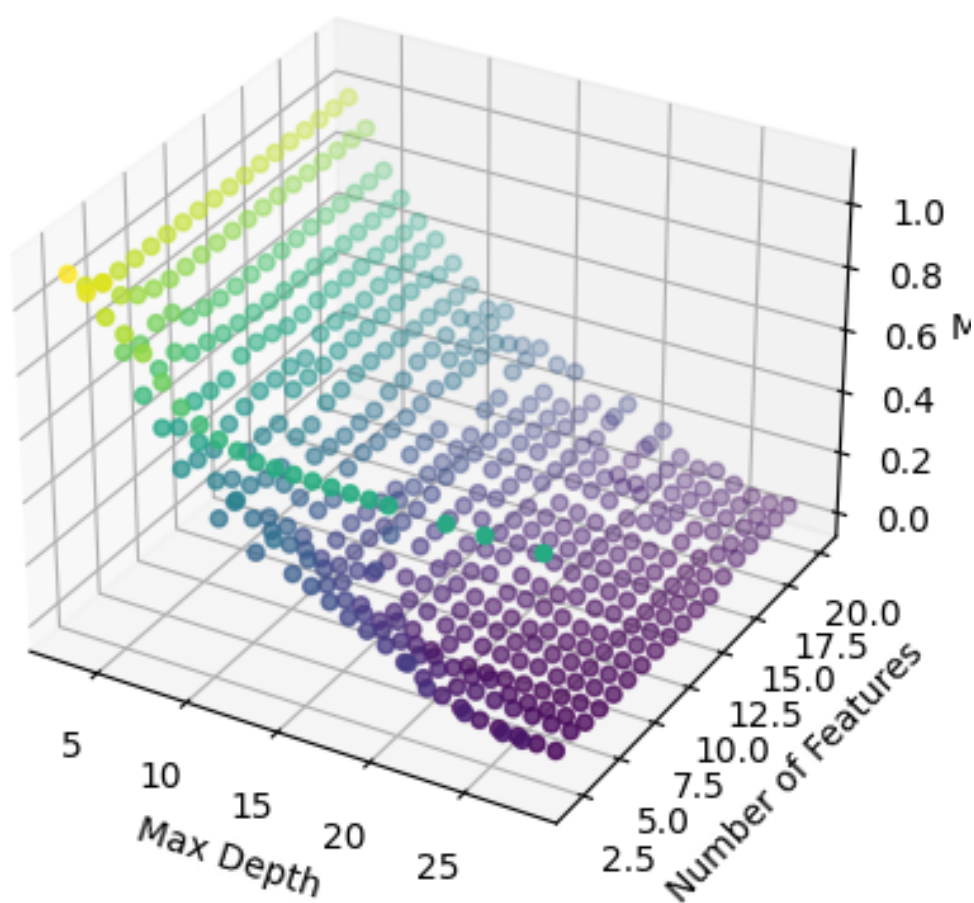
```
max_depth: 20, n_features: 13, mse_train: 89643306022.6,  
mse_test: 879912454221.0 <-
```

**Przedstawienie wyników** Wyniki przedstawiliśmy na wykresach, wykorzystując pythonowe biblioteki matplotlib.

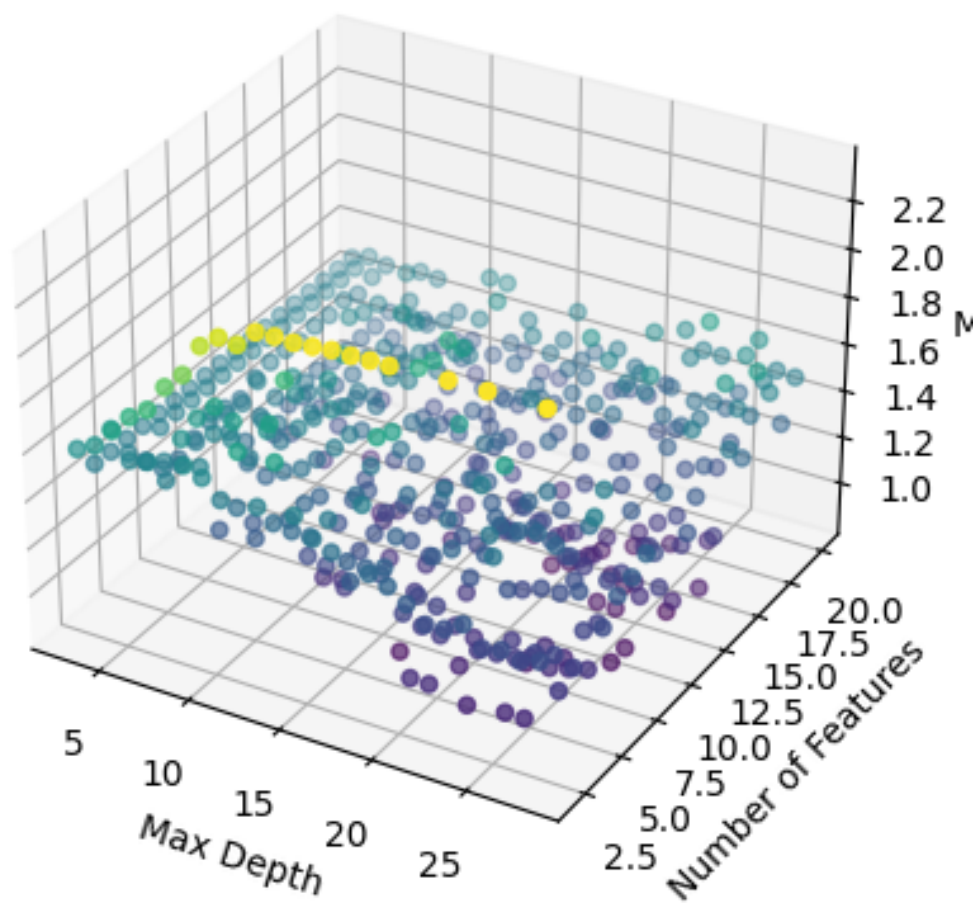


## 5 Wyniki

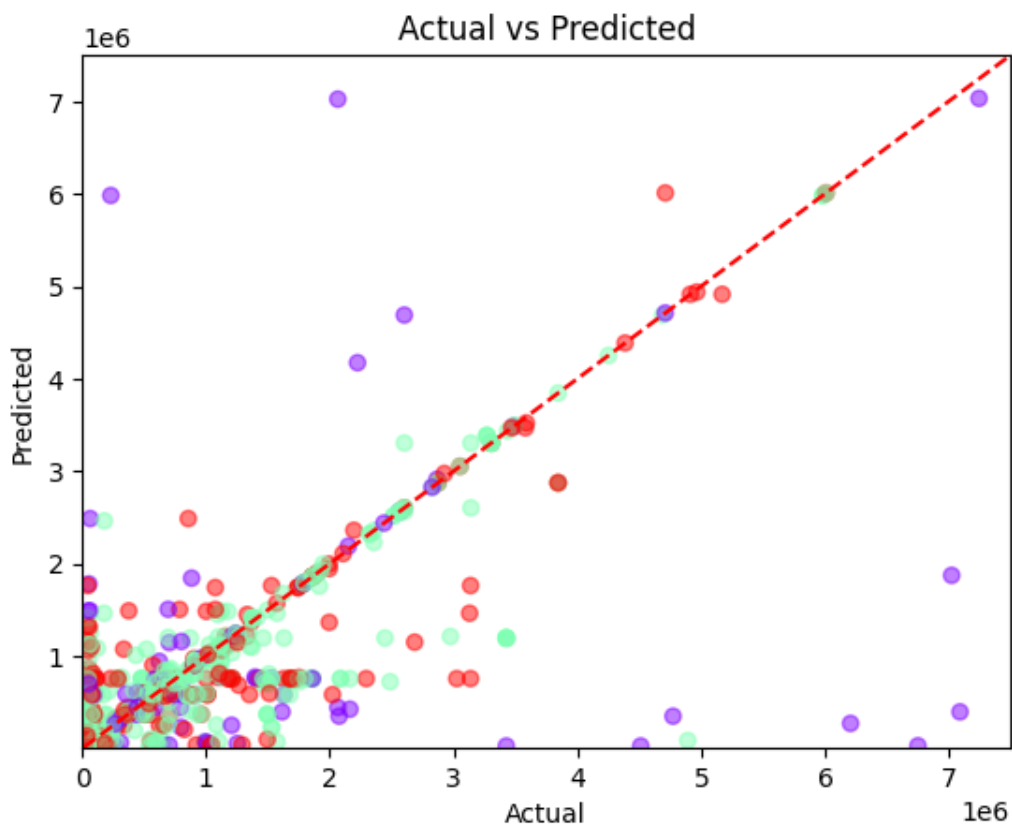
Rysunek 5: B<sub>d</sub> dla danych treningowych jako funkcja głębokości i liczby cech



Rysunek 6: B<sub>d</sub> dla danych testowych jako funkcja głębokości i liczby cech



Rysunek 7: Funkcja predykcji modelu co do wielkości finansowania porównana do prawdziwego finansowania, czerwiesze kolory odpowiadaj wikszej gstoci zaludnienia



Parametry poniej miay najwikszy zwizek z wysokoci wpywów z Unii Europejskiej do gminy.

Kategoria	Warto
Dochody z podatku od nieruchomości	0.3853
Dochody z podatku od rodków transportowych	0.2161
Powierzchnia	0.0911
Wynagrodzenie ogóem	0.0670
Dochody z podatku PCC	0.0581
Dochody razem	0.0424
Dochody z majątku	0.0292
Dochody z podatku od spadków	0.0286
Dochody z podatku rolnego	0.0277
Dochody z podatku od dziaalnoci gospodarczej	0.0225
Wynagrodzenie w relacji do redniej	0.0156
Dochody z podatku odrbne ustawy	0.0107
Dochody z podatku lenego	0.0057

Tabela 1: Najistotniejsze dane gminy wpywajce na przyznanie funduszy UE

## 6 Dyskusja

**Sukcesy** Udao nam si zebra dane z GUS-u i poczy je z danymi o inwestycjach Unii Europejskich. Stworzylimy model, który na podstawie przygotowanych przez nas danych spróbowo wykaza, jakie parametry gminy najbardziej wpyway na przyznanie rodków unijnych.

**Weryfikacja hipotezy** Nasza hipoteza zgodnie z wynikami, które uzyskalimy, okazała si **faszywa**. Nasz model za najwaniejsz dan o gminie wpywajc na przyznanie rodków z Unii Europejskiej uzna **dochód z podatków od nieruchomości**, a nie gsto zaludnienia.

**Niskie wartoci korelacji** Niestety wartoci powizania danych o gminie i wpywów z UE w naszym modelu maj niskie wartoci, najwysze rzdu 0.4, po czym drastycznie spadaj do poziomu 0.01, 0.005.

**Brak moliwoci predykcji** Nasz model **nie nadaje si** do wykorzystania w celu przewidywania wpywów inwestycji z UE do gminy w przyszoci. Wynika to z dynamicznie zmieniajcej si sytuacji geopolitycznej. W ostatnich latach

zdecydowany wpływ na działania Unii Europejskiej miały takie wydarzenia jak pandemia COVID-19 lub wojna w Ukrainie. Niemoliwe do przewidzenia wydarzenia na arenie międzynarodowej sprawiają, że predykcja przyszłych zachowań tak dużych instytucji jak Unia Europejska jest dla naszego modelu zadaniem nieosiągalnym.

## 7 Konkluzja

Analiza obala nasz hipotezę, że gęstość zaludnienia odgrywa największą rolę i zamiast tego wskazuje na dochód z podatków od nieruchomości. Nasz model, mimo że zidentyfikował pewne zależności, charakteryzuje się niskimi wartościami korelacji i ograniczoną zdolnością do przewidywania przyszłych funduszy.

Aby poprawić dokładność przyszłych analiz, sugerujemy wykorzystanie innych technik modelowania (gradient boosting, sieci neuronowe) oraz dodatkowych zmiennych, takich jak zmiany polityczne, ekonomiczne i społeczne. Rozważenie tych dynamicznych czynników może lepiej odzwierciedlić skomplikowane procesy decyzyjne w Unii Europejskiej i zwiększyć trafność prognozowania przyznawania funduszy.

## Literatura

- [1] Hendrik Blockeel, Laurens Devos, Benoît Frénay, Géraldine Nanfack, and Siegfried Nijssen. Decision trees: from efficient prediction to responsible ai. *Frontiers in Artificial Intelligence*, 6, 2023.
- [2] Riccardo Crescenzi, Marco Di Cataldo, and Mara Giua. It's not about the money. eu funds, local opportunities, and euroscepticism. *Regional Science and Urban Economics*, 84:103556, 2020.
- [3] Ministerstwo Finansów. Szacunkowe dane o wykonaniu budżetu państwa w 2023 r., 2024. Accessed: 2024-05-04.
- [4] money.pl. Tak Polska skorzysta na członkostwie w UE. Zyskamy setki miliardów złotych, 2024. Accessed: 2024-05-04.
- [5] Polska Akademia Nauk. 20 lat Polski w UE, 2024. Accessed: 2024-05-04.

- [6] Marko ostar, Vladimir Ristanovi, and Chamaru de Alwis. Application of successful eu funds absorption models to sustainable regional development. *Economies*, 11(9), 2023.