

# Analiza statystyczna przyznawania funduszy UE gminom

Krzysztof Rudnicki, Michał Sar

June 6, 2024

## Contents

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

## Abstract

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

## 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, wpłacając 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. Połączyć 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 pracą, 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 łączyliśmy 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 jedną 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 wykorzystał ankiety, pytając 244 osób o to, jak efektywnie wykorzystywane były fundusze UE w Polsce, Słowenii, Węgrzech 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 przydał się nam w ocenie, jakie parametry pozytywnie wpływają na korzystanie z funduszy UE i jakie moglibyśmy śledzić 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 Walię z uwagi na różnicę w ilości pieniędzy przekazanych poszczególnym rejonom. 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 [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.

Figure 1: Strona GUS z zaznaczonymi popularnymi podgrupami

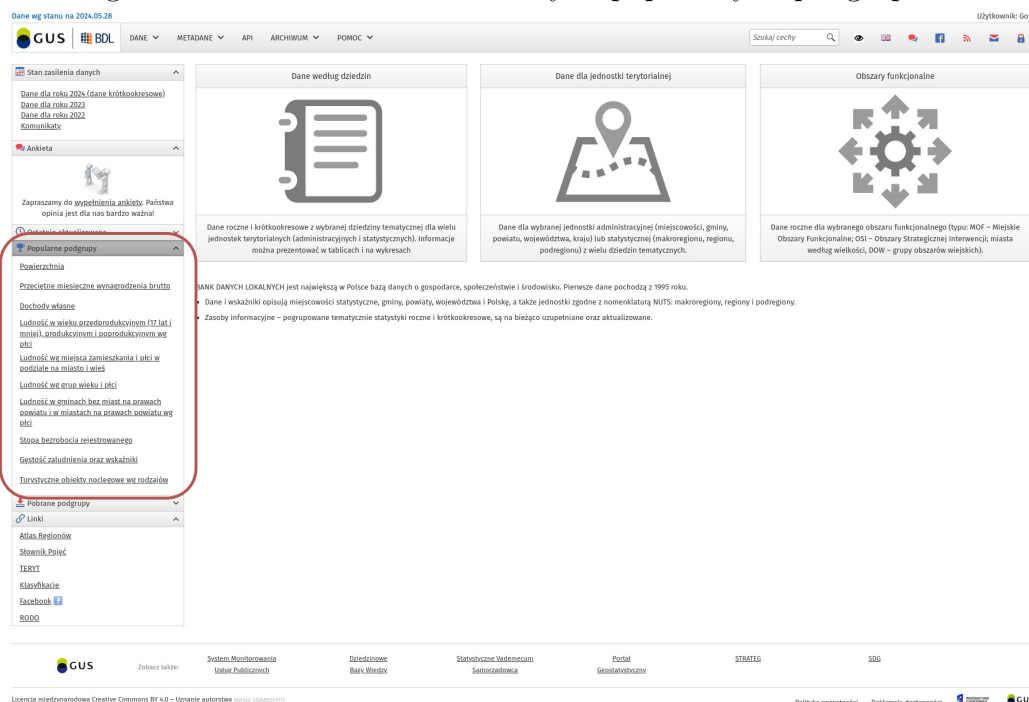







Figure 2: Strona GUS z zaznaczonymi latami i powierzchnią



Dane wg stanu na 2024.05.28 Użytkownik: Gość

  DANE ▾ METADANE ▾ API ARCHIWUM ▾ POMOC ▾   

Start / Dane według dziedzin / 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

Zaznaczonych: 8/29 ☒ ☐

**Powierzchnia**

ogółem w ha
ogółem w km2

Zaznaczonych: 1/2 ☒ ☐

[Dalej](#) [Pobierz](#)

Figure 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)

Dzierżoniów (1)

Pieszycy (3)

Pieszycy - miasto (4)

Pieszycy - obszar wiejski (5)

Piława Górna (1)

Dzierżoniów (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)

Dzierżoniów (1)

Pieszycy (3)

Pieszycy - miasto (4)

Pieszycy - obszar wiejski (5)

Piława Górna (1)

Dzierżoniów (2)

Elementów do wyboru: 4309

Wybranych elementów: 4309

→ Dalej

Figure 4: Dane o powierzchni z możliwością eksportu do CSV

Wstatistika aktualizacja 03.04.2024

Tablica Wykres Mapa

Wybór jednostek terytorialnych Agregaty Kod Puste Export Objaśnienia

Jednostka terytorialna ▲

POLSKA		2017	2018
		[km2]	[km2]
	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

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

**Przygotowywanie danych** Dane musieliśmy przeprosować 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 płęć, wiek przedprodukcyjny/produkcyjny/poprodukcyjny, gęstość zaludnienia).
3. Województwo.
4. Wymeldowania i zameldowania.

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

Dzieliśmy 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 wykazywał 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 pythonową bibliotekę matplotlib.

## 5 Wyniki

Figure 5: Błąd dla danych treningowych jako funkcja głębokości i liczby cech

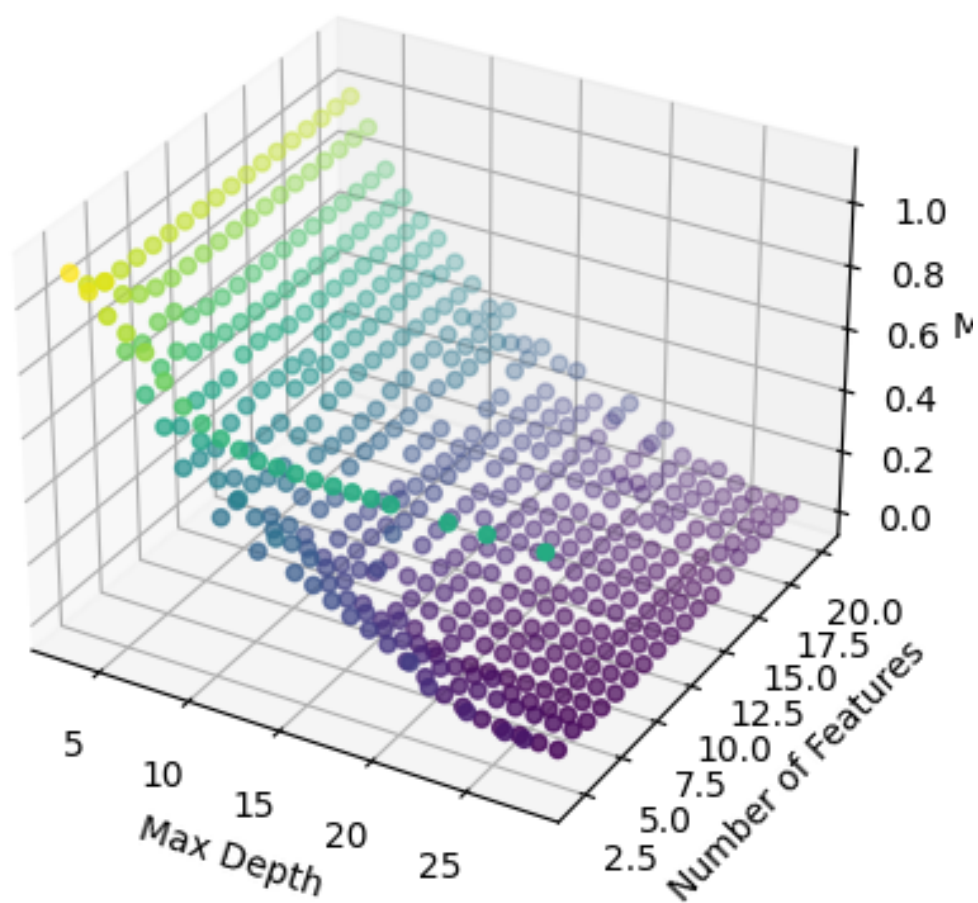


Figure 6: Błąd dla danych testowych jako funkcja głębokości i liczby cech

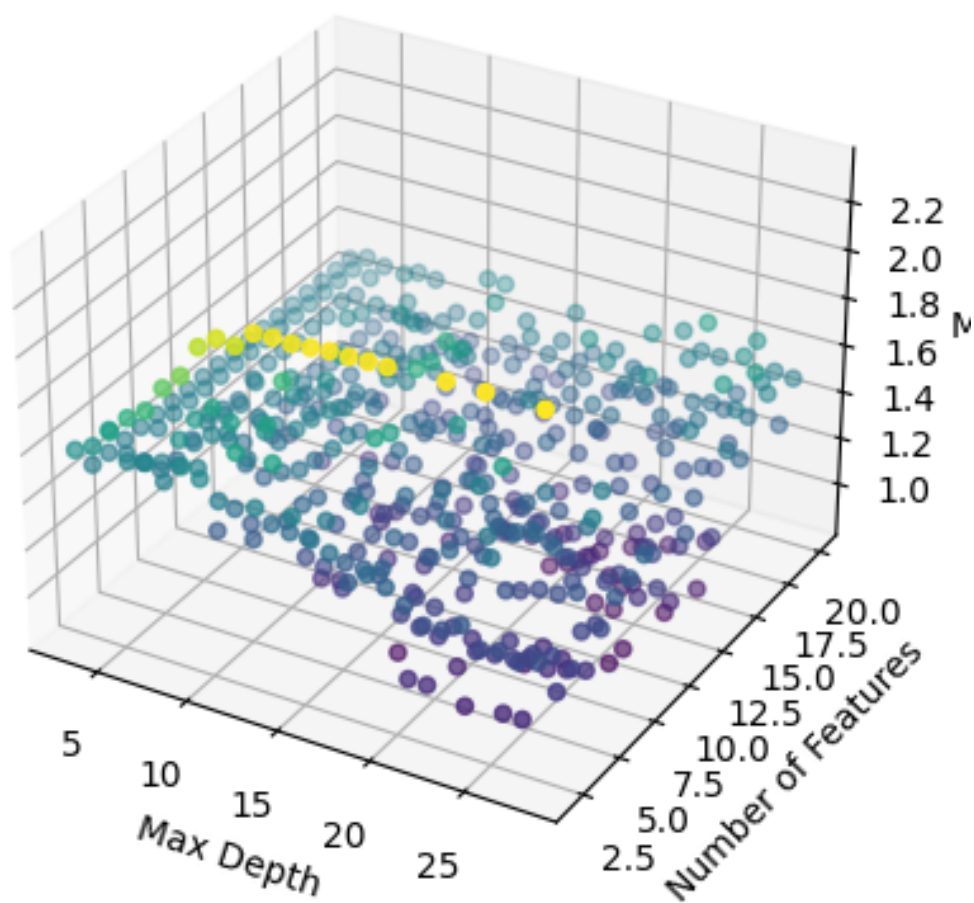
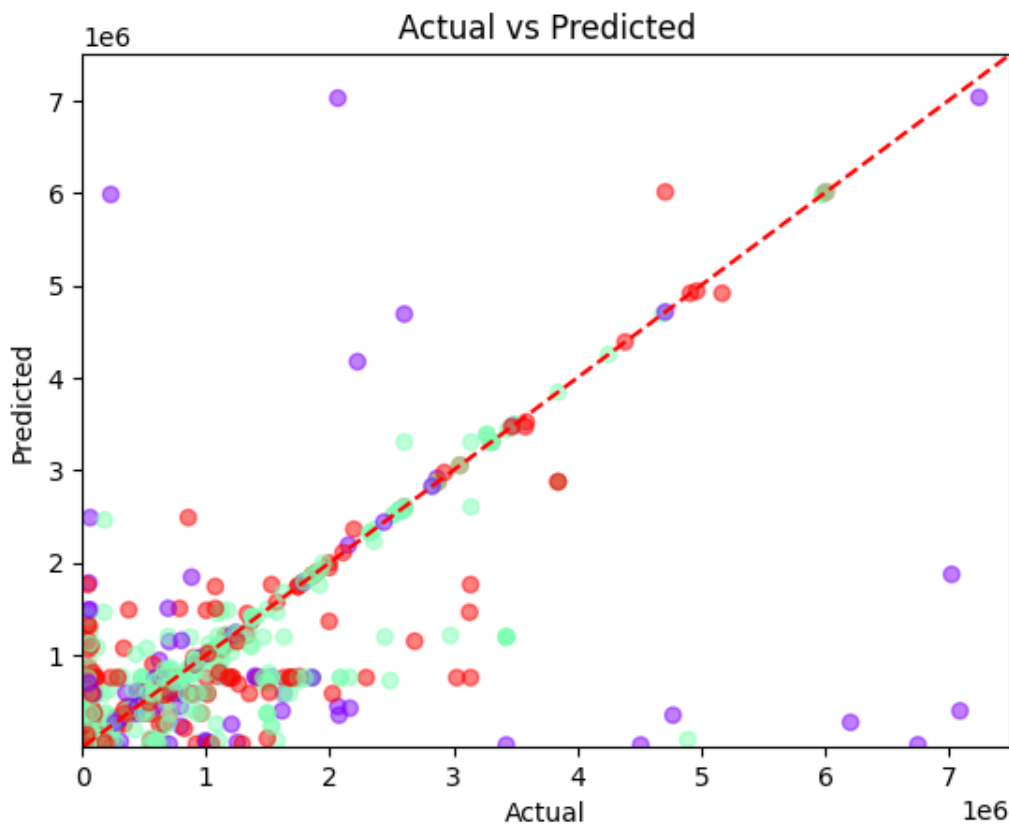


Figure 7: Funkcja predykcji modelu co do wielkości finansowania porównana do prawdziwego finansowania, czerwienie kolory odpowiadają większej gęstości zaludnienia



Parametry poniżej miały największy związek z wysokością wpływów z Unii Europejskiej do gminy.

Kategoria	Wartość
Dochody z podatku od nieruchomości	0.3853
Dochody z podatku od środków transportowych	0.2161
Powierzchnia	0.0911
Wynagrodzenie ogółem	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 działalności gospodarczej	0.0225
Wynagrodzenie w relacji do średniej	0.0156
Dochody z podatku odrębne ustawy	0.0107
Dochody z podatku leśnego	0.0057

Table 1: Najistotniejsze dane gminy wpływające na przyznanie funduszy UE

## 6 Dyskusja

**Sukcesy** Udało nam się zebrać dane z GUS-u i połączyć je z danymi o inwestycjach Unii Europejskich. Stworzyliśmy model, który na podstawie przygotowanych przez nas danych spróbował wykazać, jakie parametry gminy najbardziej wpływały na przyznanie środków unijnych.

**Weryfikacja hipotezy** Nasza hipoteza zgodnie z wynikami, które uzyskaliśmy, okazała się **fałszywa**. Nasz model za najważniejszą daną o gminie wpływającą na przyznanie środków z Unii Europejskiej uznał **dochód z podatków od nieruchomości**, a nie gęstość zaludnienia.

**Niskie wartości korelacji** Niestety wartości powiązania danych o gminie i wpływów z UE w naszym modelu mają niskie wartości, najwyższe rzędu 0.4, po czym drastycznie spadają do poziomu 0.01, 0.005.

**Brak możliwości predykcji** Nasz model **nie nadaje się** do wykorzystania w celu przewidywania wpływów inwestycji z UE do gminy w przyszłości.

Wynika to z dynamicznie zmieniającej się sytuacji geopolitycznej. W ostatnich latach zdecydowany wpływ na działania Unii Europejskiej miały takie wydarzenia jak pandemia COVID-19 lub wojna w Ukrainie. Niemożliwe 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 obaliła naszą 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.

## References

- [1] Hendrik Blockeel, Laurens Devos, Benoît Frénay, Géraldine Nanack, 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 skorzystała na członkostwie w UE. Zyskałiśmy setki miliardów złotych, 2024. Accessed: 2024-05-04.
- [5] Polska Akademia Nauk. 20 lat Polski w UE, 2024. Accessed: 2024-05-04.

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