

Dokumentacja

Optymalizacja zbioru publikacji pod kątem maksymalizacji punktów: algorytm zachłanny, referencyjny

Ewelina Chmielewska

Mateusz Nosek

Cel projektu:

Celem projektu było zaimplementowanie w języku Python algorytmu zachłannego będącego referencyjnym, którego celem była maksymalizacja liczby punktów uzyskiwanych przez uczelnię za zgłaszane publikacje oraz sprawdzenie jego wydajności i poprawności na zbiorze testowym.

Opis ogólny problemu:

Za publikacje – artykuły, monografie (autorstwo, redakcję, rozdział), artykuły recenzyjne, przekłady oraz edycję źródeł (autorstwo, redakcję, rozdział) – przyznaje się określoną ilość punktów. Punkty te są następnie dzielone pomiędzy autorów, zgłaszających swój wkład w dzieło. Liczba punktów, jaką można przyznać danemu autorowi ogółem, albo za konkretne dzieło jest jednak ograniczona i zależy między innymi od czasu pracy i udziału w dyscyplinie danego autora, a także od typu stworzonego przez niego dzieła. Autorzy mogą upoważniać jednostkę do swoich publikacji, lecz nałożone na nich osobno ograniczenia pozostają w mocy. Liczba dzieł zgłaszanych przez jednostkę również podlega pewnym limitom, zależnym od liczby zatrudnianych przez nią pracowników oraz ich działalności. Deklarowanie przez pracowników dyscyplin, upoważnianie jednostki do swoich publikacji oraz deklarowanie przynależności do N w zatrudniających podmiotach zwiększają ogólne ograniczenie liczby dzieł.

Naszym zadaniem jest zaimplementowanie algorytmu zachłannego referencyjnego, który wybierając dzieła do publikacji przez jednostkę będzie przestrzegał ograniczeń wynikających ze specyfikacji problemu oraz będzie maksymalizował punkty uzyskiwane przez jednostkę. Projekt w uproszczeniu można sprowadzić do dwóch problemów plecakowych – jeden z nich wynika z ograniczeń narzuconych na autora publikacji, zaś drugi z ograniczeń narzuconych na jednostkę.

Modyfikacja problemu i algorytmu:

W projekcie zastosowano uproszczoną wersję problemu: zmianie uległy ograniczenia nałożone zarówno na autora, jak i na jednostkę. Opis nowych w dalszej części dokumentacji.

Do standardowego działania algorytmu zachłannego również wprowadzone zostały pewne modyfikacje. Po zwróceniu listy zaakceptowanych publikacji, każda z nich może zostać usunięta z listy z pewnym prawdopodobieństwem alfa. W kolejnej iteracji dopełniamy dotychczasowe rozwiązanie publikacjami, które ostatecznie nie weszły w jego skład.

Reprezentacja rozwiązania

Rozwiązanie reprezentowane jest przez macierz, w której kolejne rzędy odpowiadają poszczególnym autorom, zaś kolumny kolejnym publikacjom.

Oznaczenia:

0 – dana publikacja nie została wliczona do rozwiązania końcowego

1 – dana publikacja została wliczona do rozwiązania końcowego

Opis działania programu:

Zaprojektowany algorytm uruchamiany jest dla grupy przypadków testowych z następującymi punktami startowymi (dla każdego z nich 28 razy):

- 1 raz – żadna publikacja żadnego autora nie wchodzi w skład początkowego rozwiązania
- 1 raz – wszystkie publikacje każdego autora wchodzi w skład początkowego rozwiązania
- 1 raz – dla każdego autora wybrane są te jego publikacje, które zostały wybrane jako pierwsze k z listy posortowanej względem stosunku liczby punktów przynależnych autorowi podzielonych przez stopień współautorstwa
- 25 razy – dla każdego autora wybrane są te jego publikacje, które zostały wybrane jako pierwsze k z listy ustawionych losowo

Opis działania programu dla pojedynczego testu:

1. Ustalenie listy wstępnie zaakceptowanych publikacji.
2. Heurystyczne wyznaczenie długości listy publikacji przeznaczonych do opublikowania HLPP.
3. Załadowanie listy autorów i ich publikacji.
Podczas ładowania publikacji zaakceptowanych, sprawdzane jest spełnienie ograniczeń. Publikacje, które oznaczone zostały jako zaakceptowane, ale nie spełniają ograniczeń pozostają na liście publikacji danego autora, ale są usuwane z listy publikacji zaakceptowanych dla tego autora.
4. Wyznaczenie liczby obliczeń funkcji celu, po których ma nastąpić zapis najlepszego znajdującego dotychczas wyniku.
5. Wyznaczenie publikacji do zaakceptowania i zapis najlepszego wyniku jeśli osiągnięto próg.
6. Jeśli warunek stopu (liczba obliczeń funkcji celu mniejsza od maksymalnej zadanej) nie jest spełniony, powrót do punktu 5.
7. Sprawdzenie spełnienia ograniczeń.
8. Zwrócenie wyniku: ostatecznej wartości funkcji celu oraz wektora reprezentującego zaakceptowane publikacje.

Algorytm:

1. Posortowanie publikacji, które nie znajdują się na liście publikacji zaakceptowanych ze względu na punkty_za_publicację / udział_w_publicacji.
2. Wyznaczenie początkowej funkcji celu: sumy punktów za publikacje zaakceptowane.
3. Uaktualnienie heurystycznej liczby publikacji HLPP, jakie pozostały do opublikowania.
4. Wyznaczenie początkowej wartości funkcji heurystycznej: sumy wartości punktów
5. Dodawanie do listy publikacji do zaakceptowania kolejnych publikacji
Dla kolejnych publikacji z listy publikacji do rozpatrzenia:
 - a) wyznaczenie heurystycznej wartości sumy punktów za pozostałe do zaakceptowania publikacje dla przypadku, jeśli rozważana publikacja zostanie odrzucona - HA
 - b) wyznaczenie heurystycznej wartości sumy punktów za pozostałe do zaakceptowania publikacje dla przypadku, jeśli rozważana publikacja zostanie zaakceptowana - HO
 - c) wyznaczenie tymczasowej funkcji celu TG: suma AG i punktów za publikację
 - d) Wybór korzystniejszego rozwiązania: AG + HO lub TG + HA
 - e) Odpowiednie uaktualnienie wartości funkcji celu i wartości funkcji heurystycznej
6. Uaktualnienie najlepszego dotychczas znajdującego wyniku
7. Wyznaczenie losowo publikacji z listy publikacji zaakceptowanych do usunięcia z tej listy.

HLPP – heurystyczna liczba publikacji, jakie pozostały do opublikowania

AG - aktualna wartość funkcji heurystycznej

Funkcja celu:

$$f = \sum_i \left(\sum_j a_{i,j} w_{i,j} \right)$$

$a_{i,j}$ - przedstawienie do ewaluacji wkładu autora i w dzieło j; może mieć wartość 0 albo 1

$w_{i,j}$ - kwant punktowej wartości przeliczeniowej autora i w dziele j

Żadne kary nie zostały uwzględnione w funkcji celu, gdyż algorytm wymaga, by każde rozwiązanie cząstkowe było poprawne. Z tego samego powodu **nie jest również konieczne występowanie funkcji naprawy**.

Pierwotne ograniczenia:

- 1. Limit zgłaszanego współautorstwa dzieł dla pracowników badawczych badawczo-dydaktycznych**

$$\sum u(n, y) z(n, y) \leq 4 \text{ udział}(n)$$

Suma przedstawionych do ewaluacji udziałów jednostkowych w dziełach dla danego autora n musi być nie większa od czterokrotności średniej wartości iloczynu wymiaru czasu pracy tego autora i jego udziału w dyscyplinie.

- 2. Limit zgłaszanego współautorstwa monografii (a także tłumaczeń i opracowań) dla pracowników badawczych i badawczo-dydaktycznych**

$$\sum u(n, y) z(n, y) \leq 2 \text{ udział}(n)$$

Suma przedstawionych do ewaluacji udziałów jednostkowych w monografiach dla danego autora n musi być nie większa od dwukrotności średniej wartości iloczynu wymiaru czasu pracy tego autora i jego udziału w dyscyplinie.

Ograniczenie to nie obejmuje doktorantów.

y – publikacja będąca autorstwem, redakcją lub rozdziałem w monografiach za nie więcej niż 100 punktów nominalnych

- 3. Limit zgłaszanego współautorstwa dla doktorantów**

$$\sum u(d, y) z(d, y) \leq 4$$

Suma udziałów jednostkowych doktoranta d w monografiach, które dla tego doktoranta przedstawione zostały do ewaluacji musi być nie większa od czterech.

- 4. Ograniczenie ogólnej liczby dzieł dla jednostki ewaluowanej**

$$\sum u(x, y) z(x, y) \leq 3N - 3N_0 - 6N_1 - 6N_2$$

Suma przedstawionych do ewaluacji udziałów jednostkowych autorów w publikacjach, do których upoważniona została jednostka nie może być większa od $3N - 3N_0 - 6N_2 - 6N_3$.

y – publikacja autora, który upoważnił jednostkę

5. Ograniczenie liczby monografii i rozdziałów dla jednostki ewaluowanej (5%)

$$\sum u(x, y) z(x, y) \leq 0.15N$$

Suma przedstawionych do ewaluacji udziałów jednostkowych autorów w monografiach, do których upoważniona została jednostka nie może być większa od 0.15N

y – publikacje będące monografiami autorskimi, ich redakcją albo rozdziałami, włączając tłumaczenia

6. Ograniczenie liczby dzieł osób spoza N dla jednostki ewaluowanej (20% dzieł)

$$\sum u(x, y) z(x, y) \leq 0.6N$$

Suma przedstawionych do ewaluacji udziałów jednostkowych doktorantów oraz osób niewłączonych do N w publikacjach, do których upoważniona została jednostka nie może być większa od 0.6N

y – publikacja autora, który upoważnił jednostkę

Oznaczenia:

d - doktorant

n – pracownik

x – autorzy, którzy upoważnili jednostkę

y – dzieło

z – przedstawienie publikacji do ewaluacji

u(x, y) – udział jednostkowy autora x w dziele y

udział(n) – średnia wartość iloczynu wymiaru czasu pacy i udziału w dyscyplinie.

1. brak drugiej dyscypliny w jednostce → udział = udział + 0.25

2. udział < 0.25 → udział = 0.25

3. udział > 1 → udział = 1

N – liczba pracowników

N₀ – liczba pracowników N, którzy nie upoważnili jednostki do żadnej publikacji

N₁ – liczba pracowników badawczych lub badawczo dydaktycznych, prowadzących działalność naukową, ale nie deklarujących przynależności do N w żadnym z zatrudniających podmiotów

N₂ – liczba pracowników badawczych lub badawczo dydaktycznych, którzy nie deklarowali przynależności do dyscypliny

Nowe ograniczenia:

Nowe ograniczenia stanowią uproszczoną wersję wcześniejszych ograniczeń.

1. Ograniczenie sumy współudziału dla każdego autora

$$\sum u(n, y) z(n, y) \leq 4$$

2. Ograniczenie sumy współudziału dla wszystkich autorów dyscypliny

$$\sum u(x, y) z(x, y) \leq 3N$$

Oznaczenia:

x – autorzy, którzy upoważnili jednostkę

y – dzieło

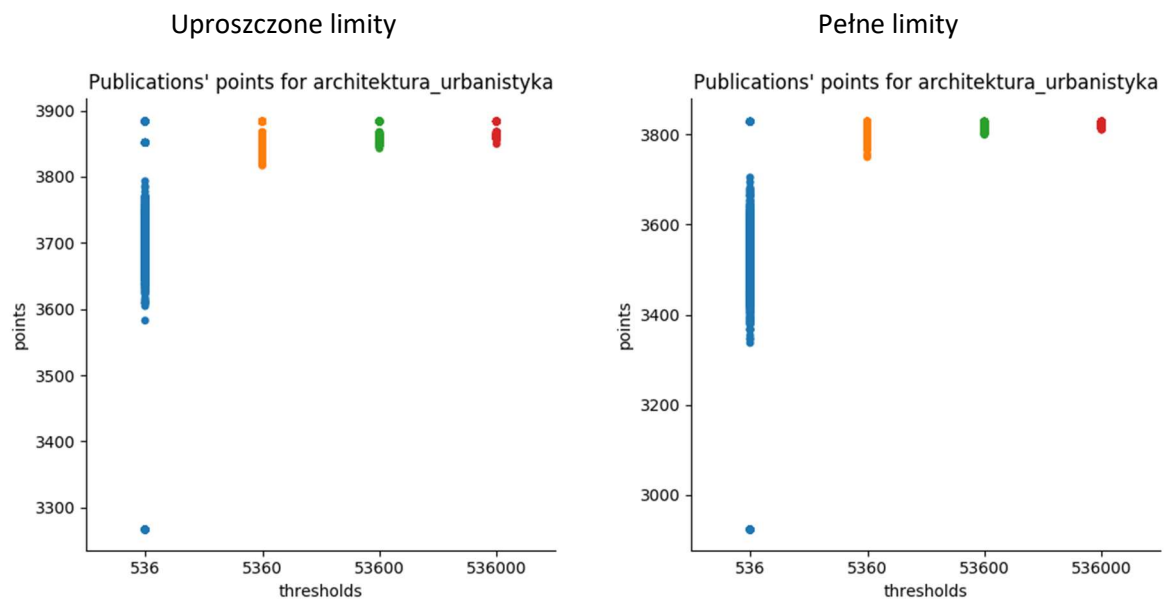
z – przedstawienie publikacji do ewaluacji

u(x, y) – udział jednostkowy autora x w dziele y

N – liczba pracowników

Wyniki:

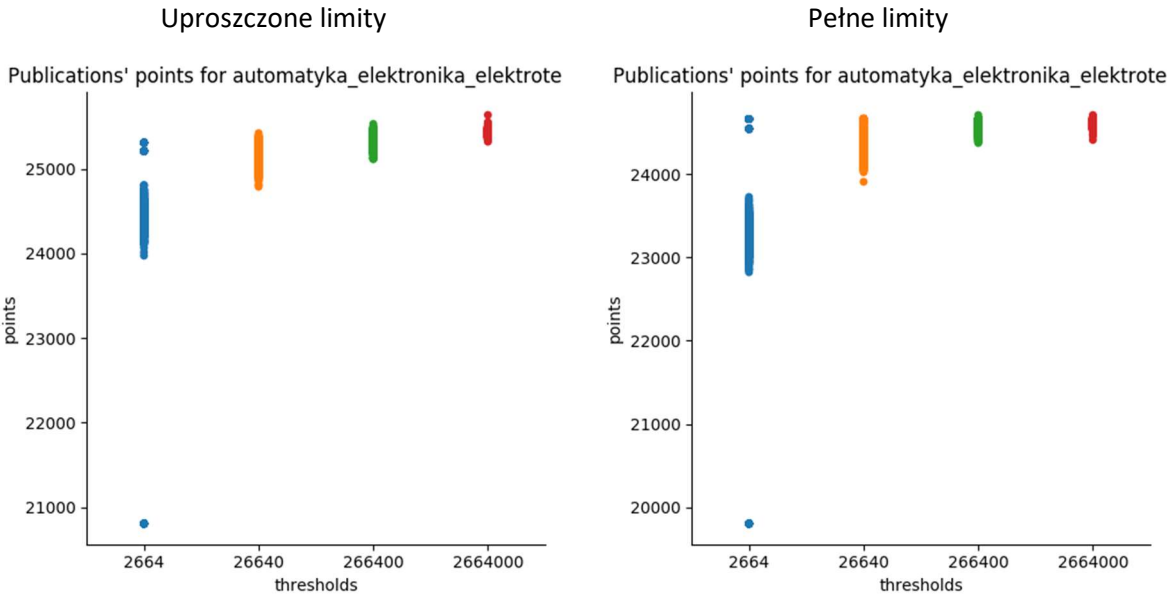
1. Architektura urbanistyka



	Uprozczone limity
Funkcja celu	3883.46
Plik	architektura_urbanistyka_0_8_0.txt

	Pełne limity
Funkcja celu	3829.185
Plik	architektura_urbanistyka_3_11_22.txt

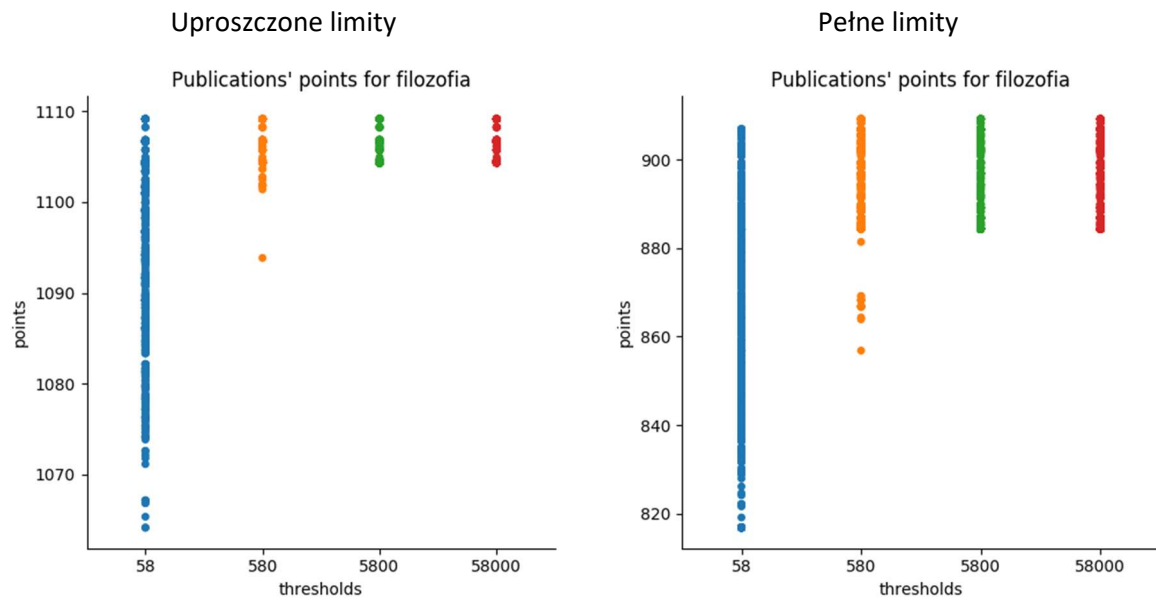
2. Automatyka elektronika elektrotechnika



	Uprozczone limity
Funkcja celu	25639.653
Plik	automatyka_elektronika_elektrotechnika_3_15_23.txt

	Pełne limity
Funkcja celu	24707.68
Plik	automatyka_elektronika_elektrotechnika_3_14_0.txt

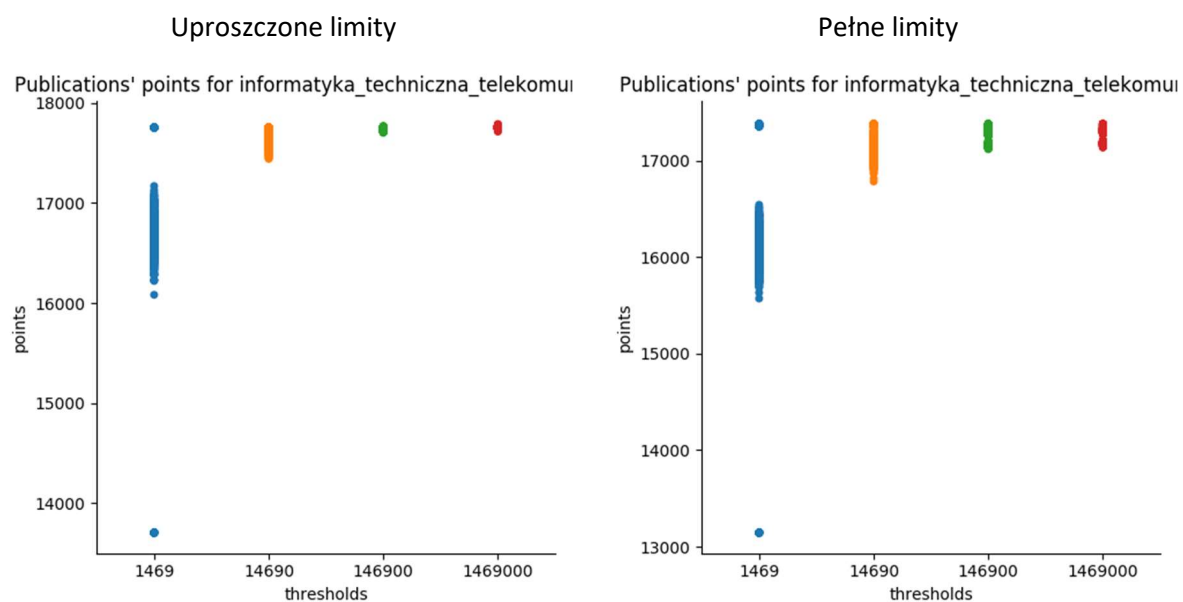
3. Filozofia



	Uproszczone limity
Funkcja celu	1109.15
Plik	filozofia_3_3_12.txt

	Pełne limity
Funkcja celu	909.15
Plik	filozofia_3_15_1.txt

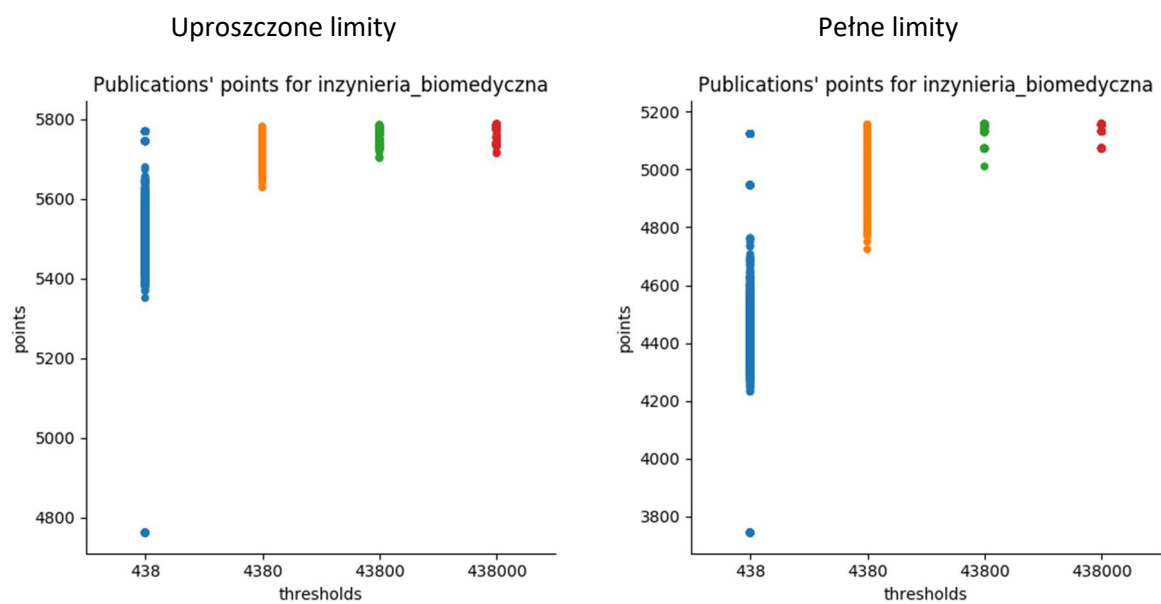
4. Informatyka techniczna telekomunikacja



	Uproszczone limity
Funkcja celu	17790.626
Plik	informatyka_techiczna_telekomunikacja_3_26_16.txt

	Pełne limity
Funkcja celu	17377.265
Plik	informatyka_techiczna_telekomunikacja_0_13_0.txt

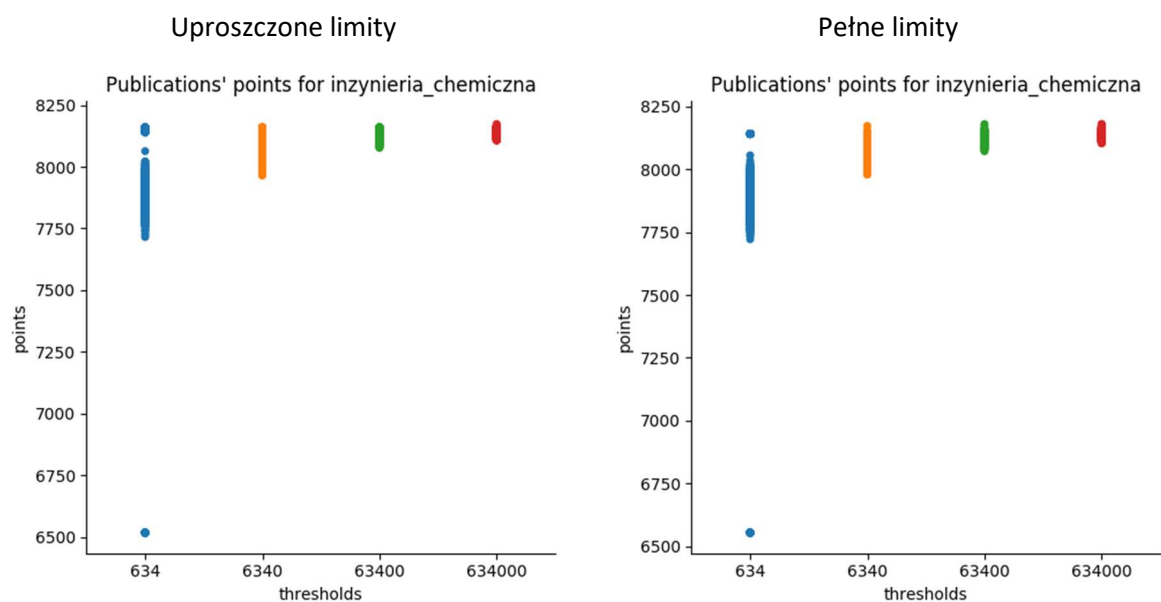
5. Inżynieria biomedyczna



	Uprozczone limity
Funkcja celu	5787.433
Plik	inzynieria_biomedyczna_3_9_14.tx

	Pełne limity
Funkcja celu	5156.697
Plik	inzynieria_biomedyczna_0_7_0.txt

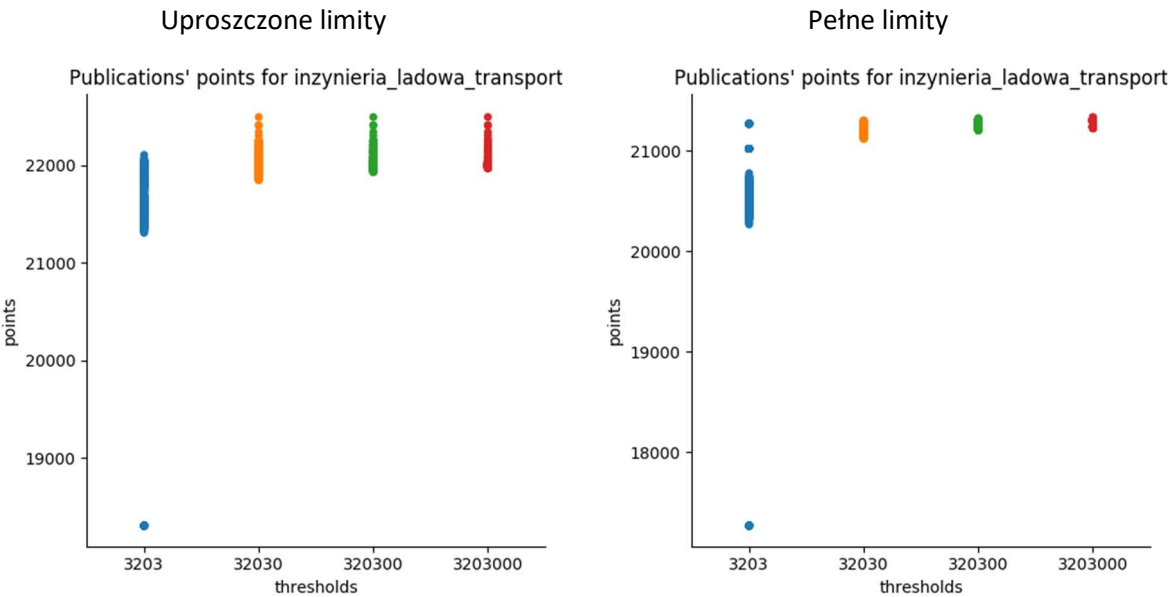
6. Inżynieria chemiczna



	Uprozczone limity
Funkcja celu	8172.933
Plik	inzynieria_chemiczna_3_16_17.txt

	Pełne limity
Funkcja celu	8180.863
Plik	inzynieria_chemiczna_3_19_0.txt

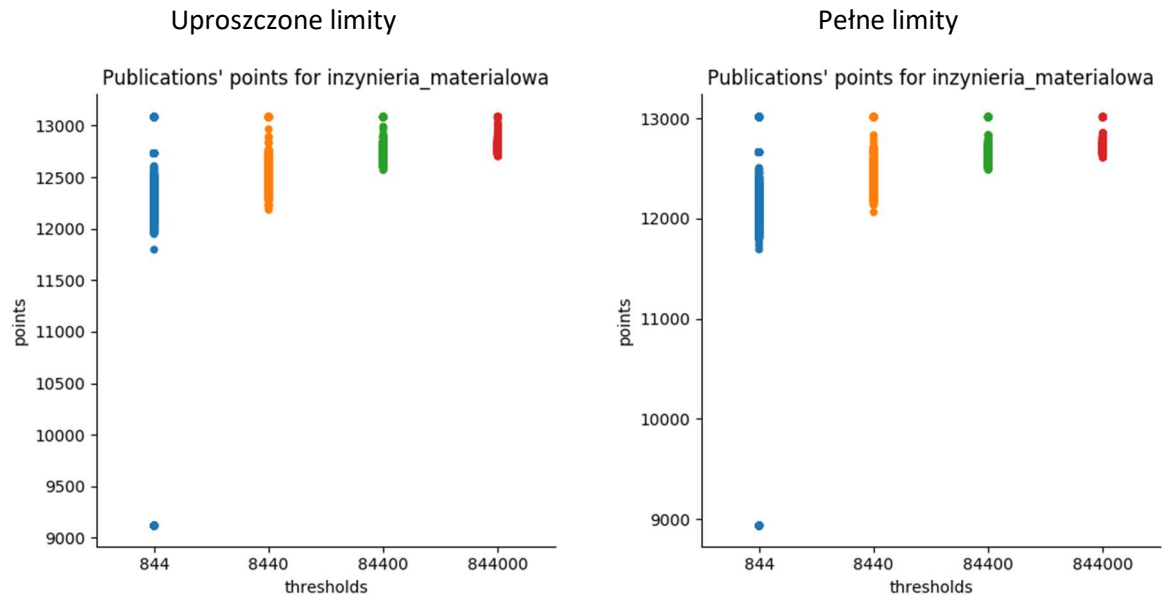
7. Inżynieria ładowa transport



	Uproszczone limity
Funkcja celu	22493.037
Plik	inzynieria_ladowa_transport_0_6_0.txt

	Pełne limity
Funkcja celu	21336.77
Plik	inzynieria_ladowa_transport_3_22_12.txt

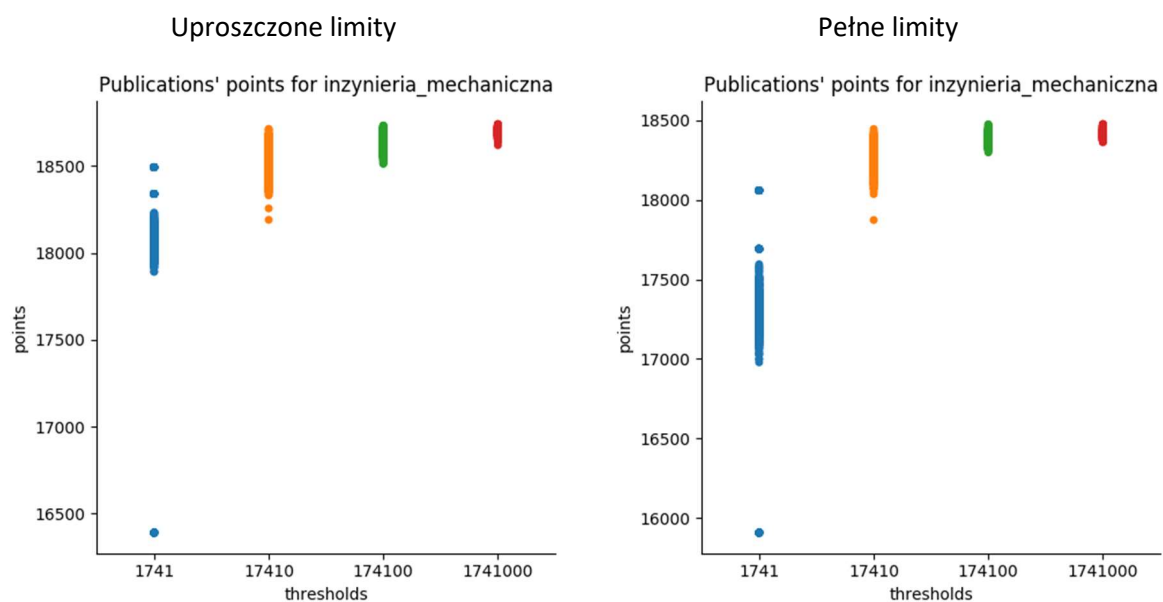
8. Inżynieria materiałowa



	Uprozczone limity
Funkcja celu	13083.395
Plik	inzynieria_materialowa_2_25_0.txt

	Pełne limity
Funkcja celu	13013.912
Plik	inzynieria_materialowa_2_25_0.txt

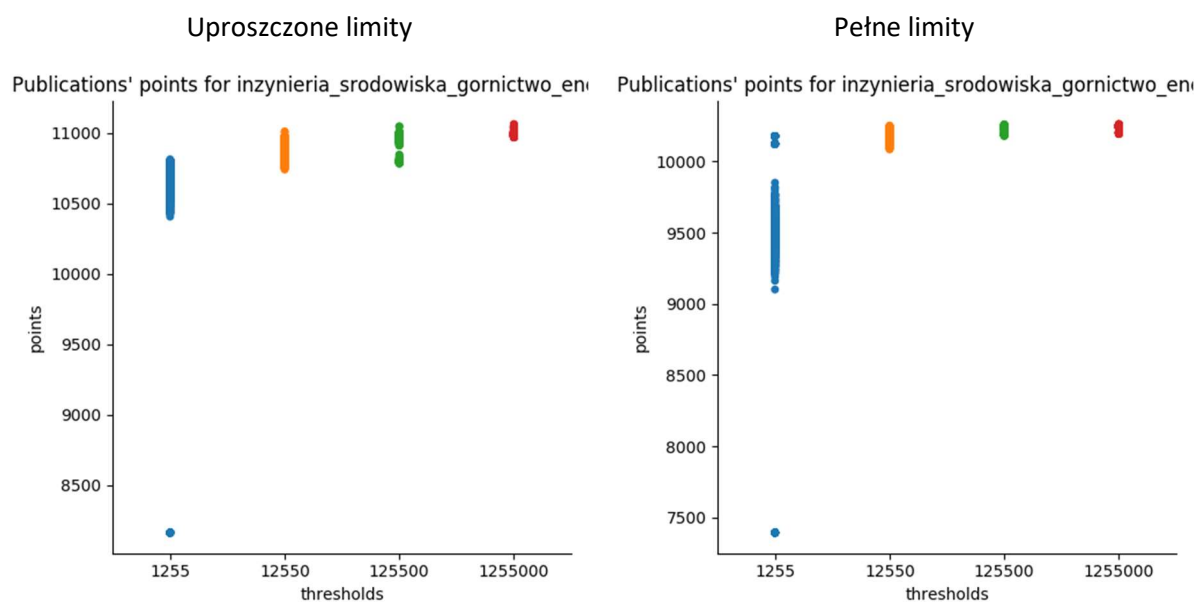
9. Inżynieria mechaniczna



	Uprozczone limity
Funkcja celu	18740.93
Plik	inzynieria_mechaniczna_3_8_7.txt

	Pełne limity
Funkcja celu	18475.661
Plik	inzynieria_mechaniczna_3_13_12.txt

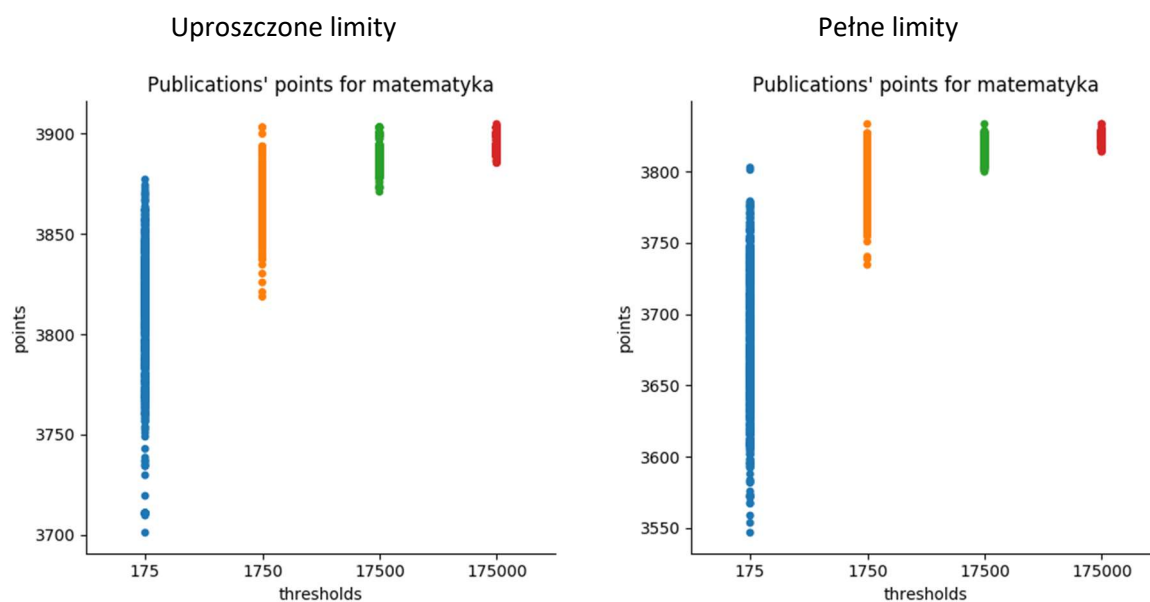
10. Inżynieria środowiska górnictwo energetyka



	Uprozczone limity
Funkcja celu	11065.164
Plik	inzynieria_srodowiska_gornictwo_energetyka_3_23_15.txt

	Pełne limity
Funkcja celu	10261.768
Plik	inzynieria_srodowiska_gornictwo_energetyka_3_6_15.txt

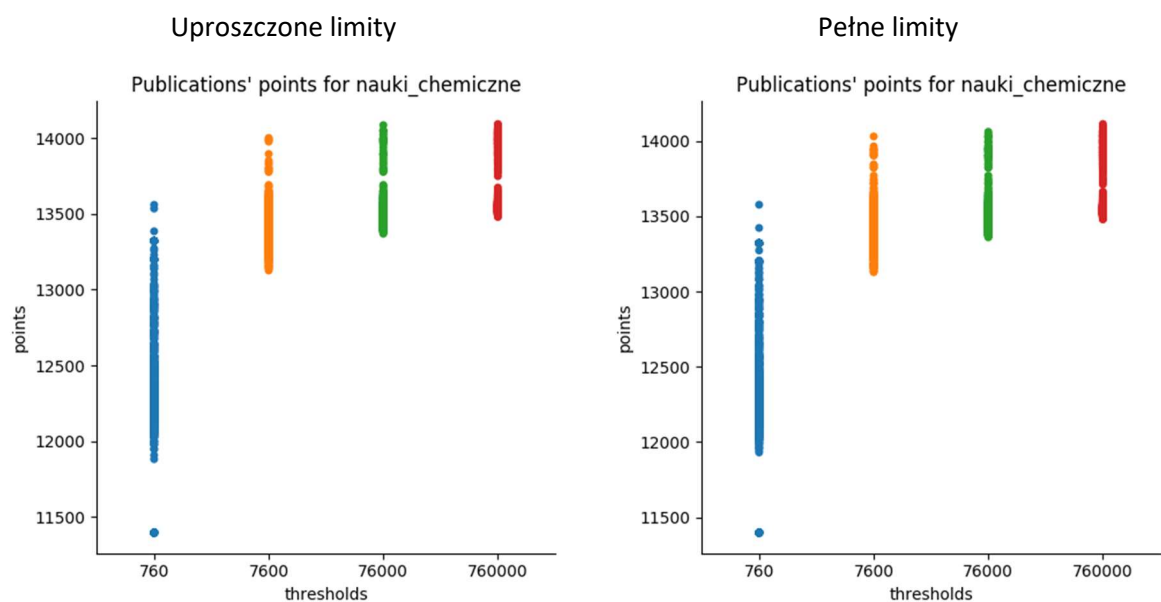
11. Matematyka



	Uproszczone limity
Funkcja celu	3904.875
Plik	matematyka_3_3_21.txt

	Pełne limity
Funkcja celu	3833.35
Plik	matematyka_3_18_14.txt

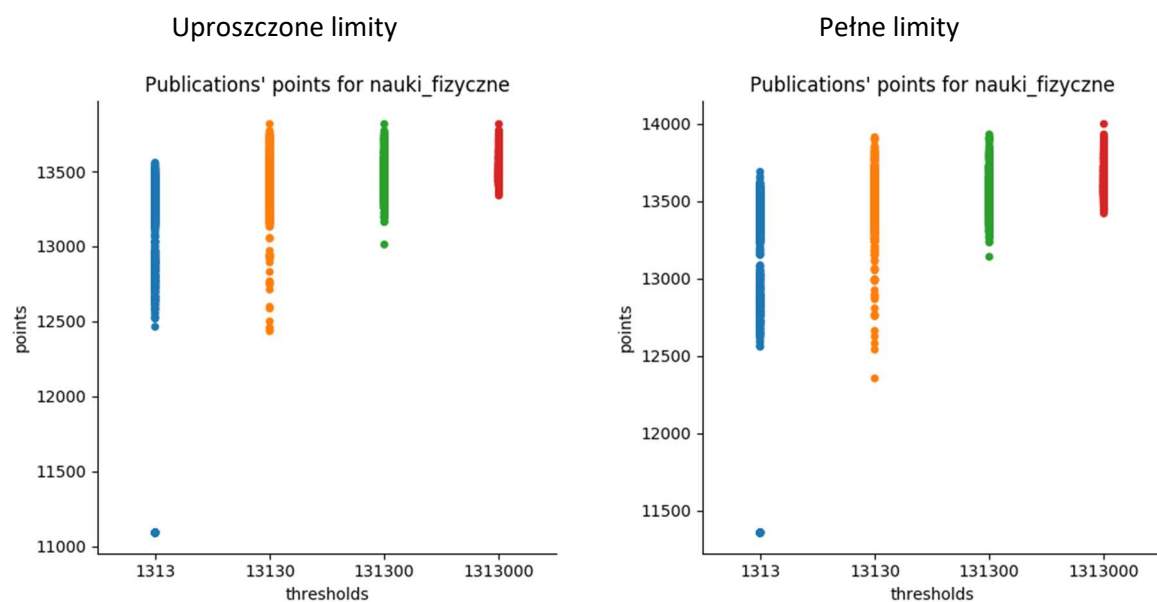
12. Nauki chemiczne



	Uproszczone limity
Funkcja celu	14091.404
Plik	nauki_chemiczne_3_2_14.txt

	Pełne limity
Funkcja celu	14113.0
Plik	nauki_chemiczne_3_20_2.txt

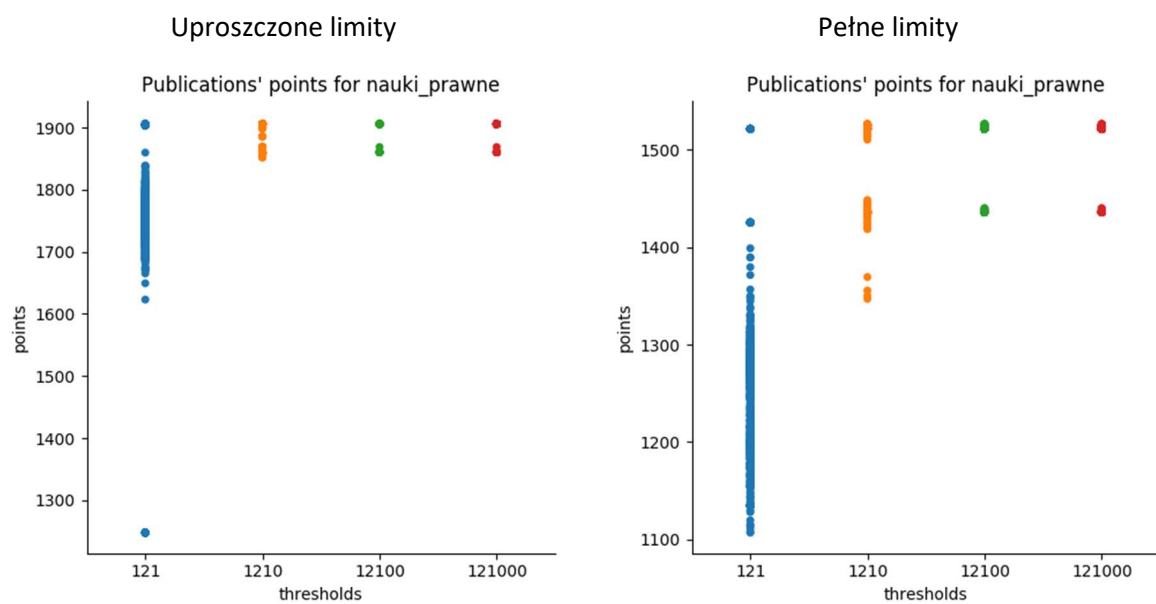
13. Nauki fizyczne



	Uproszczone limity
Funkcja celu	13815.746
Plik	nauki_fizyczne_0_5_0.txt

	Pełne limity
Funkcja celu	13997.137
Plik	nauki_fizyczne_1_19_0.txt

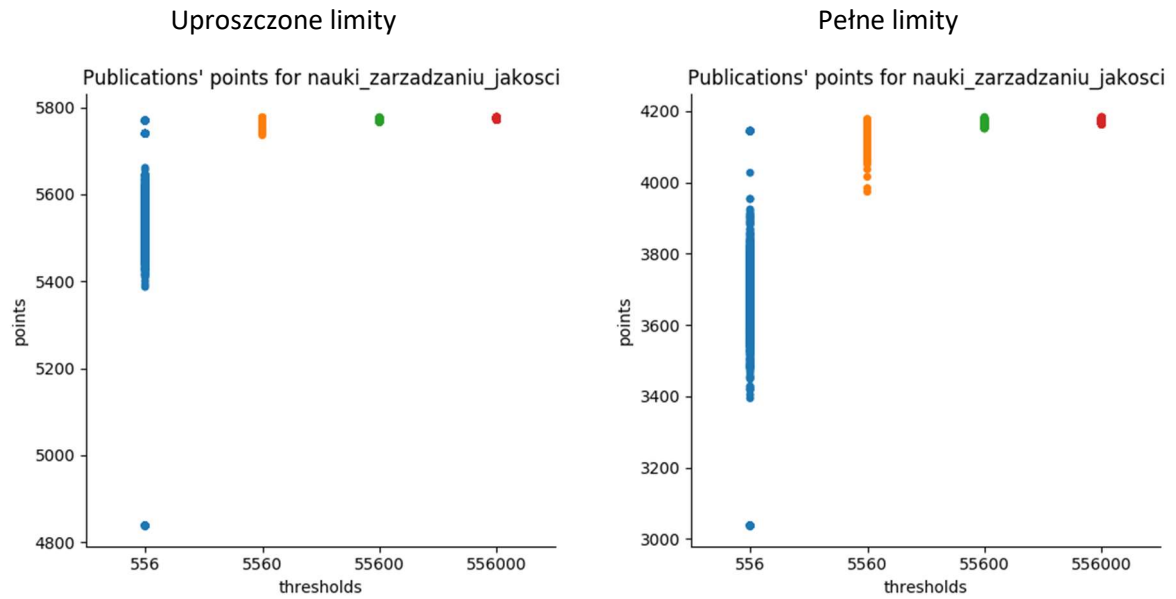
14. Nauki prawne



	Uproszczone limity
Funkcja celu	1905.6
Plik	nauki_prawne_3_1_24.txt

	Pełne limity
Funkcja celu	1526.4
Plik	nauki_prawne_3_1_24.txt

15. Nauki zarządzania jakością



	Uproszczone limity
Funkcja celu	5777.72
Plik	nauki_zarzadzaniu_jakosci_3_11_10.txt

	Pełne limity
Funkcja celu	4183.157
Plik	nauki_zarzadzaniu_jakosci_3_19_18.txt