

KNR RoboBudex

Jakub Warkocki, Marcel Zawadzki, Miłosz Piecha

30 listopada 2025

Spis treści

1 Wstęp	3
2 Identyfikacja Problemu	3
3 Zastosowane Algorytmy oraz Zbiory Danych	4
3.1 Zastosowane Zbiory Danych	4
3.2 Zastosowane Algorytmy	7
4 Finalny Produkt	7

1 Wstęp

Plik stanowi dokumentację projektu zespołu KNRRoboBudex na hackaton **Best Hacking League** w roku 2025 w konkurencji **AI**. Zadaniem konkursowym było stworzenie rozwiązania redukujące szkody środowiskowe związane z zapotrzebowaniem energetycznym oraz managementem energii cieplnej zaawansowanej infrastruktury informatycznej m.in. Data Centers. Zespół KNRRoboBudex wystąpił w trzyosobowym składzie:

- **Jakub Warkocki** (Kapitan)
- **Marcel Zawadzki**
- **Miłosz Piecha**

Dokumentacja zostanie podzielona na trzy etapy:

- **Identyfikacja Problemu** - część dokumentująca zrozumienie problematyki postawionego zadania oraz rozumowanie zespołu na temat rozwiązania realnego problemu z jego dziedziny.
- **Zastosowane Algorytmy oraz Zbiory Danych** - część prezentująca wybrane przez zespół techniki sztucznej inteligencji oraz zbiory danych rozwiązujące postawiony problem.
- **Finalny Produkt** - część opisująca zintegrowane rozwiązanie dla klienta.
- **Wnioski Ogólne** - część podsumowująca całokształt wykonanej pracy oraz zdobytej wiedzy podczas hackatonu.

Celem dokumentacji jest kompleksowe przybliżenie czytelnikowi wyzwania z którym mierzył się zespół wraz z prezentacją zaproponowanego produktu oraz rozwiązaniem problemu.

2 Identyfikacja Problemu

Rozbudowa centrów danych w celu rozszerzania usług cloud i AI zwiększa zużycie energii elektrycznej i emisje ciepła, które pozostaje niewykorzystane i zmarnowane, a w przypadku chłodzenia wodą może zakłócać ekosystemy wodne - podgrzana woda uwolniona bezpośrednio do środowiska wpływa na procesy metaboliczne i zmniejsza ilość rozpuszczonego tlenu.

Emitowane ciepło można jednak odpowiednio wykorzystać. W przypadku dużych Data Center, jest ono skupione w jednym miejscu, często daleko od potencjalnych odbiorców ciepła. Ten problem można rozwiązać wykorzystując wiele małych centrów danych do ogrzewania obiektów będących większymi odbiorcami ciepła lub normalnie samemu je wytwarzających, takich jak szkoły, zwłaszcza w zimnych obszarach gdzie chłodzenie pochłania mniej energii a ogrzewanie jest bardziej pożąданie. Ponadto, można zlokalizować je w regionach, gdzie warunki klimatyczne sprzyjają efektywnej produkcji energii z OZE, których rozbudowa mogłaby rekompensować ślad węglowy nowo powstałych klastrów obliczeniowych.

Realizowaną usługą będzie dobór optymalnej lokalizacji centrów danych, aby zminimalizować ślad węglowy centrów danych, jednocześnie wykorzystując produkowane ciepło. Odbiorcy ciepła, u których umieszczone zostanie centrum danych zyskają tańsze ogrzewanie, płacąc za jedynie część energii pobieranej przez małe centrum danych. Dostawca usług również zyska niższe koszty dzielonej energii elektrycznej. Ponadto seryjna produkcja małych centrów danych może być tańsza dzięki powtarzalnemu, zoptymalizowanemu projektowi. Firma również czerpie zyski z dostarczania ekologicznego rozwiązania, co może doprowadzić do preferowania jej przez potencjalnych klientów i ograniczyć obciążenie podatkowe za emisje. Małe centra danych emitują energię rzędu kilkuset kW, co pokrywa się z zapotrzebowaniem placówki takiej jak średniej wielkości szkoły lub urzędu miasta.

3 Zastosowane Algorytmy oraz Zbiory Danych

3.1 Zastosowane Zbiory Danych

Na zakres projektu określono obszar terytorium Stanów Zjednoczonych. Wybór ten podyktowany był dobrą dostępnością danych oraz celem ograniczenia wielkości dużych zbiorów GIS w przypadku map całego świata. Nic nie stoi jednak na przeszkodzie by rozszerzyć zakres na inne terytoria. Ponadto teren Stanów Zjednoczonych cechuje się dużym zróżnicowaniem warunków klimatycznych, pozwalając na dobrą wizualizację potencjału systemu. Do realizacji projektu wykorzystano następujące wartości z rastrowych danych geograficznych:

1. **Wind power density** - Średnia roczna siła wiatru określana pod względem produkcji energii elektrycznej w W/m². Dane pochodzą ze zbioru [Global Wind Atlas](#). Ten parametr bezpośrednio przekłada się na korzystne warunki do budowy elektrowni wiatrowych w danym obszarze.
2. **Solar power output** - Średnie roczne nasłonecznienie określane pod względem produkcji energii elektrycznej w kWh/m². Dane pochodzą ze zbioru [Global Solar Atlas](#). Ten parametr bezpośrednio przekłada się na korzystne warunki do budowy elektrowni fotowoltaicznych w danym obszarze.
3. **Fiber optic internet cable access** - Odległość od najbliższych istotnych magistrali i węzłów światłowodowych. Dane pochodzą ze zbioru [bbmaps od Infrastructure Connectivity Map](#). Ze względu na brak dostarczanego API lub bazy do pobrania przez twórców zbioru, wykorzystano przekonwertowany obraz mapy wyświetlającej światłowody i główne node'y światłowodów. Zastosowano rozmycie obrazu by uzyskać gradient od odległości od infrastruktury, następnie wykorzystano oprogramowanie [QGIS](#) do odwzorowania geograficznego obrazu. Uzyskano reprezentację oddającą odległość do infrastruktury sieciowej.
4. **Temperature** - Średnia roczna temperatura, również ze zbioru [Global Solar Atlas](#). Niska temperatura ma znaczenie przy doborze lokalizacji, ponieważ wiąże się zarówno z większym zapotrzebowaniem na ogrzewanie, jak i mniejszym kosztem energetycznym chłodzenia.
5. **Population density** - Lokalna gęstość populacji, pozyskana ze zbioru [HDX](#). Większa gęstość populacji oznacza większe zapotrzebowanie na ogrzewanie i możliwość delegowania usług rozproszonych z nowo powstałych centrów danych do położonych blisko skupisk ludności

Do wczytywania danych geograficznych wykorzystano bibliotekę Rasterio, wczytując dane geograficzne w formacie **tif**. Ze względu na duży rozmiar zbiorów pliki były odczytywane fragmentami do pamięci RAM. Po zestawieniu ze sobą wszystkich warstw rastrowych zostały wybrane równomiernie rozłożone losowe punkty z zapisanymi wektorami danych, które pozwalały na dalsze tworzenie modelu.

Podstawowy charakter dla systemu miały zbiory Temperatury i Populacji, określając potencjał dla wykorzystania centrum danych dla ogrzewania.

Zaobserwowano, że obszary miejskie, na których znajdują się placówki publiczne są szczególnie predysponowane do realizacji usług, szczególnie w zimniejszych regionach.

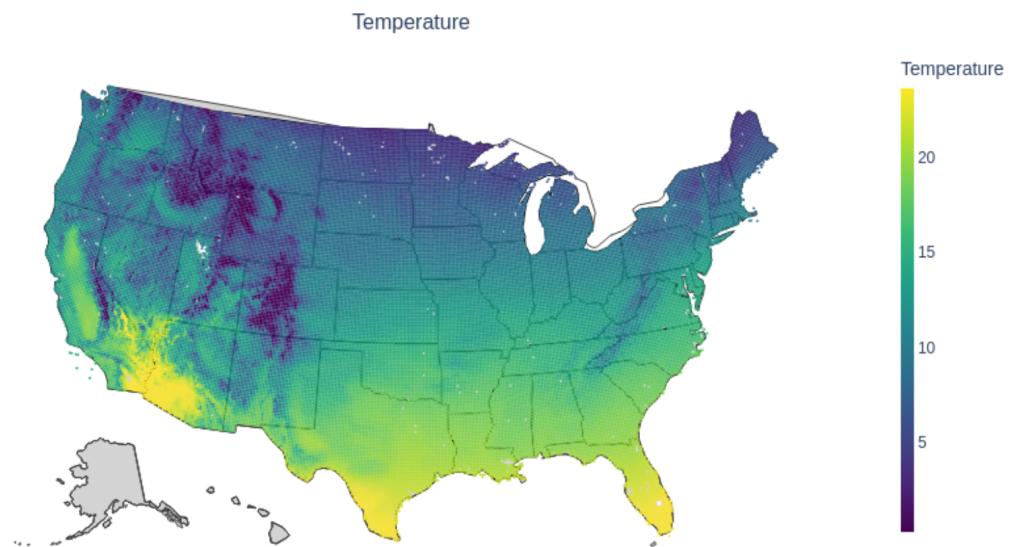
Analiza pozyskanych danych i znaczenie parametrów były kluczowe dla realizacji projektu. Możliwość produkcji energii odnawialnej, w szczególności ze słońca i wiatru, sprawia że dany obszar jest dobrze predysponowany do budowy data centre, gdyż dodany pobór przez centrum danych zostaje wytworzony ekologicznie.

Zaobserwowano, że zarówno intensywne wiatry (tereny przy zbiornikach wodnych i odpowiednie miejsca w górach) jak i duże nasłonecznienie (głównie interior USA) nadają się bardzo dobrze do realizacji instalacji OZE.

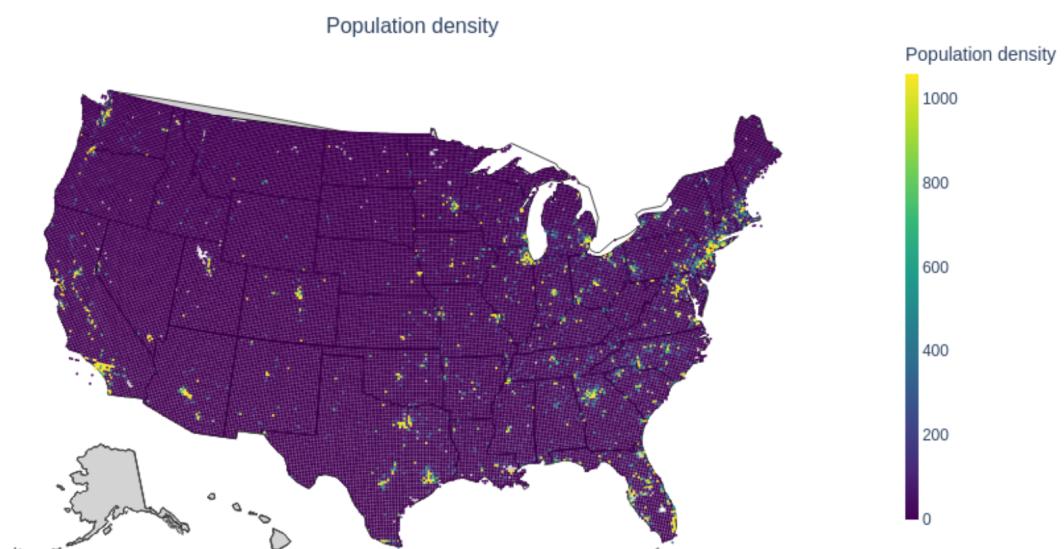
Kluczowa dla redukcji kosztów instalacji jest bliskość do instalacji światłowodowych, obserowana głównie we wschodniej części USA.

Należy się spodziewać że obszary jednocześnie: bliskie infrastruktury sieciowej, dostarczające możliwości co najmniej 1 z metod produkcji energii odnawialnej, oraz odpowiednio predysponowane klimatycznie i populacyjnie będą najlepsze do budowy proponowanych rozproszonych Data Center.

W zbiorach danych występowały dane które musiały zostać odrzucone: tereny poza granicami USA, morza i jeziora. Poza tym dane cechowały się wysoką jakością. Pod względem statystycznym, dane miały oczekiwany rozkład. Pewnym wyjątkiem były wartości średnich prędkości wiatru, których rozkład wykazał duże odchylenia standardowe. Wynika to z naturalnych warunków i wpływu górnego i dolnego na przepływ powietrza. Wartości odpowiednio znormalizowano.

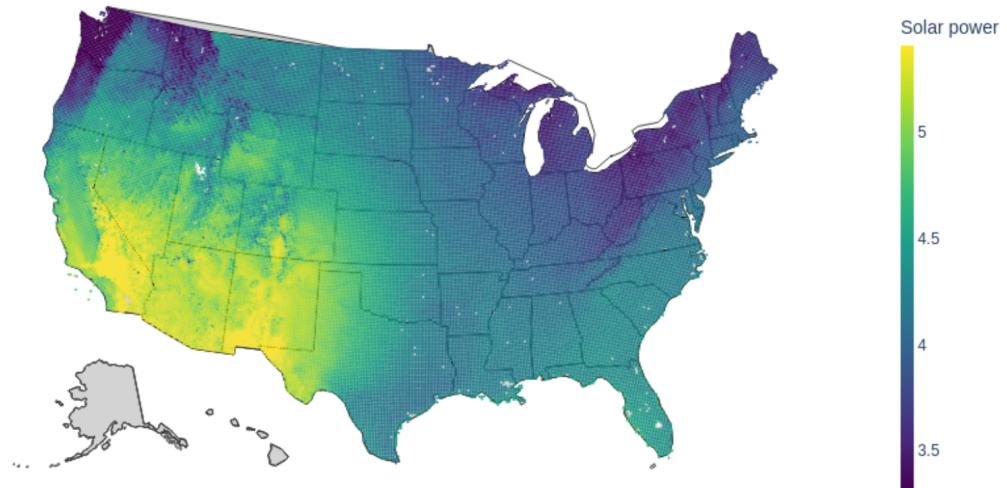


Rysunek 1: Mapa średniej temperatury.



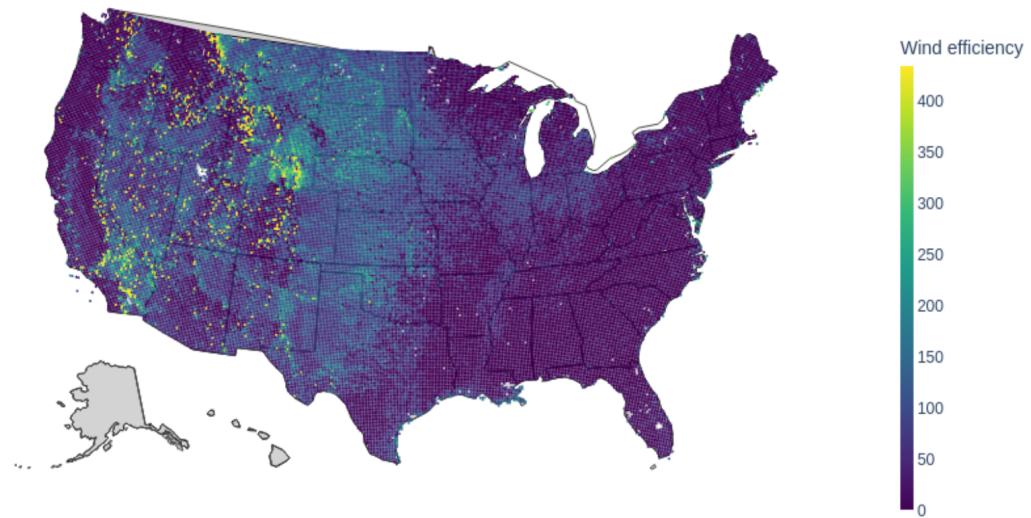
Rysunek 2: Mapa gęstości zaludnienia.

Solar power

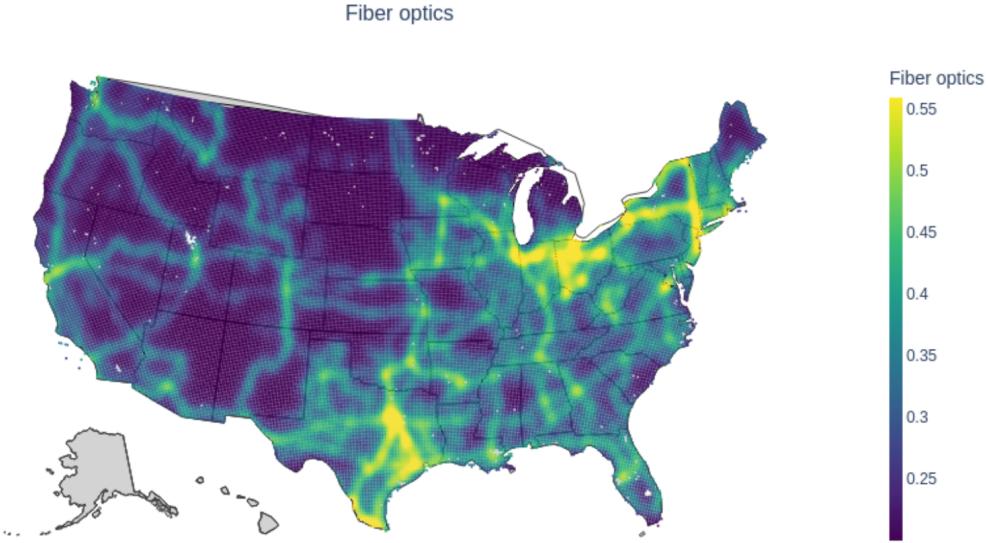


Rysunek 3: Mapa nasłonecznienia.

Wind efficiency



Rysunek 4: Mapa mocy wiatru.



Rysunek 5: Mapa dostępności do światłowodów.

Parametr	avg	σ
Wind power	78.933	344.454
Solar Power	4.364	0.499
Fiber optics	0.302	0.088
Temperature	11.698	5.347
Population density	49.616	348.032

3.2 Zastosowane Algorytmy

Rozwiązywanie postawionego problemu koncentruje się wokół stworzenia systemu inteligentnej oceny warunków środowiskowych oraz aglomeracyjnych w celu optymalnego ekologicznie oraz kosztowo rozmieszczenia centrów danych. To skomplikowane zadanie może zostać przeformułowane jako: **zdefiniowanie modelu optymalności danej lokalizacji geograficznej**. Wykorzystanie ręcznego wyboru funkcji heurystyki nie jest najlepszym rozwiązaniem, gdyż w przypadku wielu atrybutów danych geograficznych funkcja ta może przybrać postać wielomodalną, trudną do jednoznacznego zdefiniowania przez człowieka na podstawie „patrzenia na dane”.

Zadanie modelowania jakości danej lokalizacji na podstawie wyszczególnionych parametrów, może zostać oddelegowane modelowi neuronowemu. Podjęte rozwiązanie zostało oparte na architekturze **RankNET**, sieci neuronowej uczonej na podstawie dwóch próbek jednocześnie, która wyznacza dla nich wynik. Poprawność rankingu pomiędzy nimi jest oceniana na podstawie wcześniej wyznaczonych zasad eksperckich. Jest to algorytm wykorzystywany, np. przy lokalizacji nowych magazynów. Zaprojektowane zasady wynikły z przeprowadzonej analizy danych. Zaprojektowano 8 zasad, obejmujących zjawiska takie jak pozytywny wpływ na ocenę wystarczająco dużej gęstości zaludnienia i niskich temperatur powietrza, aby uzyskać potencjał odbioru ciepła. Pozytywny wpływ mają m.in. warunki do budowy instalacji energetycznej wiatowej lub fotowoltaicznej. Ponadto uwzględniono zjawiska takie jak kluczowe znaczenie potencjału odbioru ciepła w porównaniu do produkcji energii - wysoki potencjał produkcji energii bez pobliskich odbiorców ciepła nie jest użyteczny.

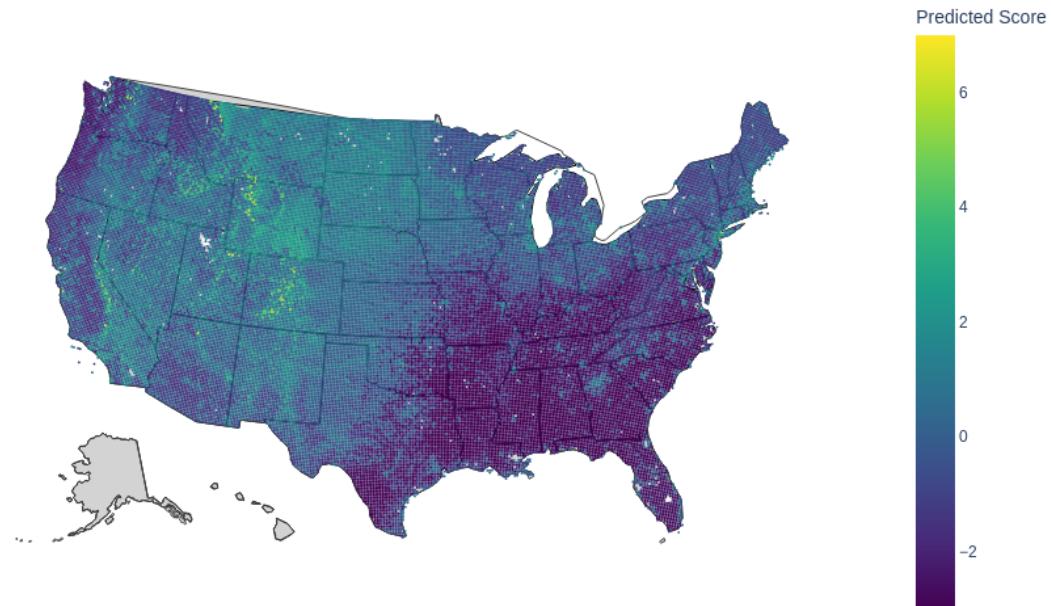
//wyznaczanie dla pojedynczego punktu / wielu lokalizacji

4 Finalny Produkt

Rozwiązywanie zaimplementowano w języku Python

Rozwiązanie pozwoliło na utworzenie mapy ze względną metryką adekwatności obszarów do lokalizacji nowych centrów danych,

Model Predictions Across USA



Rysunek 6: Mapa wyników.

Analizując otrzymaną mapę ogólną można zaobserwować liczne korzystne zjawiska. Obszary południowo-wschodnie są zdecydowanie niekorzystne - wysokie temperatury powietrza umożliwiają odbiór ciepła, a potencjał energii szczególnie wiatrowej jest niski.

Zdecydowanie lepiej wypada sytuacja na północy USA, gdzie duże wiatry i niskie temperatury są na tyle korzystne, że jednaczesne położenie blisko miast i światłowodów tworzy idealne lokalizacje na rozproszone centra danych.

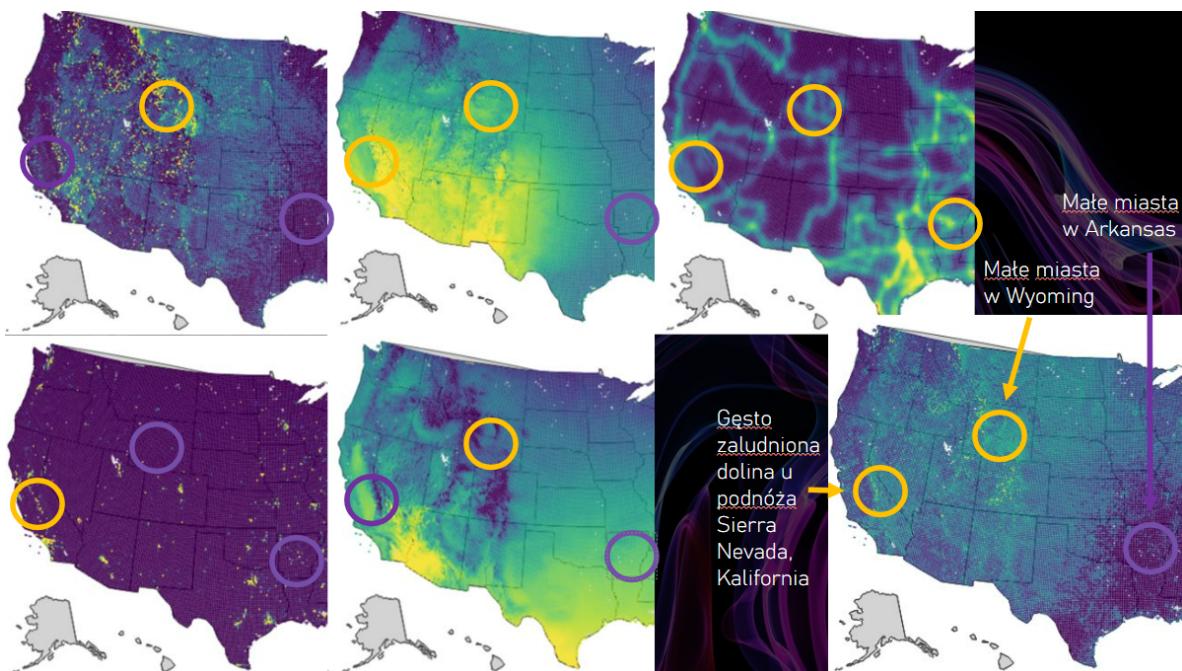
Stosunkowo dobrze wypada również środkowy zachód, gdzie jest bardzo wysoki potencjał produkcji energii ze słońca, przy nie tak wysokich temperaturach powietrza (tereny górskie).

Przykładowe tereny i zależności przedstawiono na rys. 9. Stany o największych obszarach odpowiednich do budowy rozproszonych data center to np. Montana, Wyoming i Kolorado, szczególnie mniejsze miejscowości położone niedaleko światłowodów. Wszystkie warunki pogodowe są tam odpowiednie.

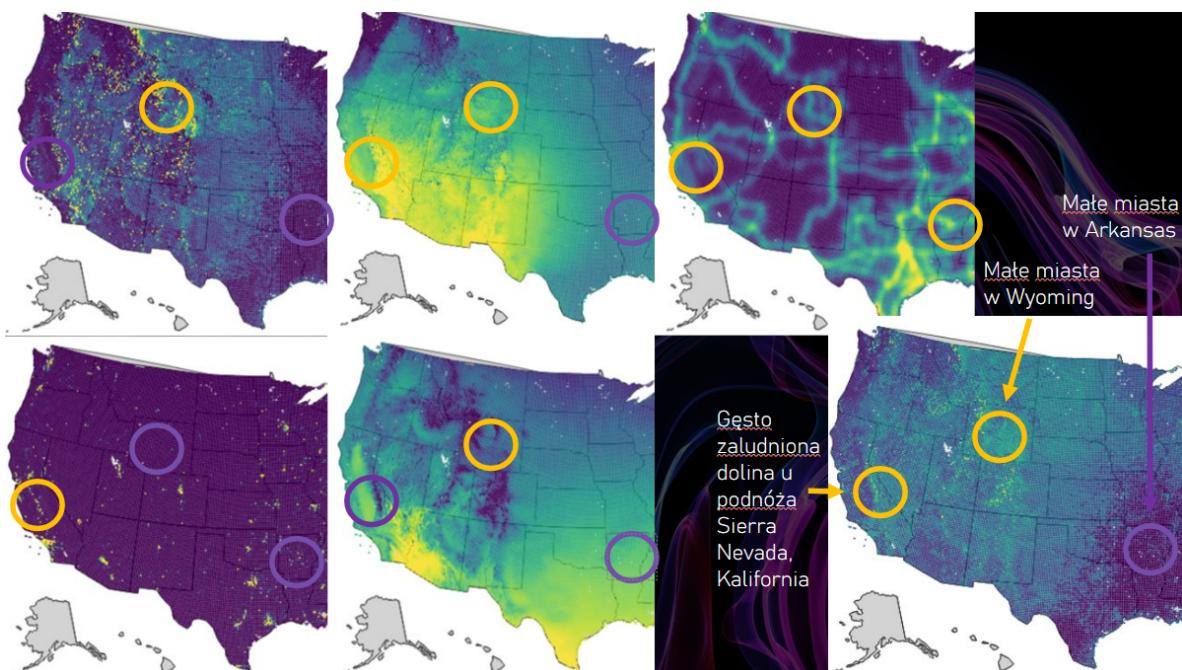
Ponadto, opracowano interfejs graficzny dla systemu, umożliwiający:

- Wyświetlanie ogólnej mapy oceny miejsc
- Wyświetlanie map wartości parametrów
- Generowanie (wybranej liczby) najlepszych lokalizacji dla centrów danych
- Dokładna ocena dla określonej współrzędnej z wysoką lokalną precyzją

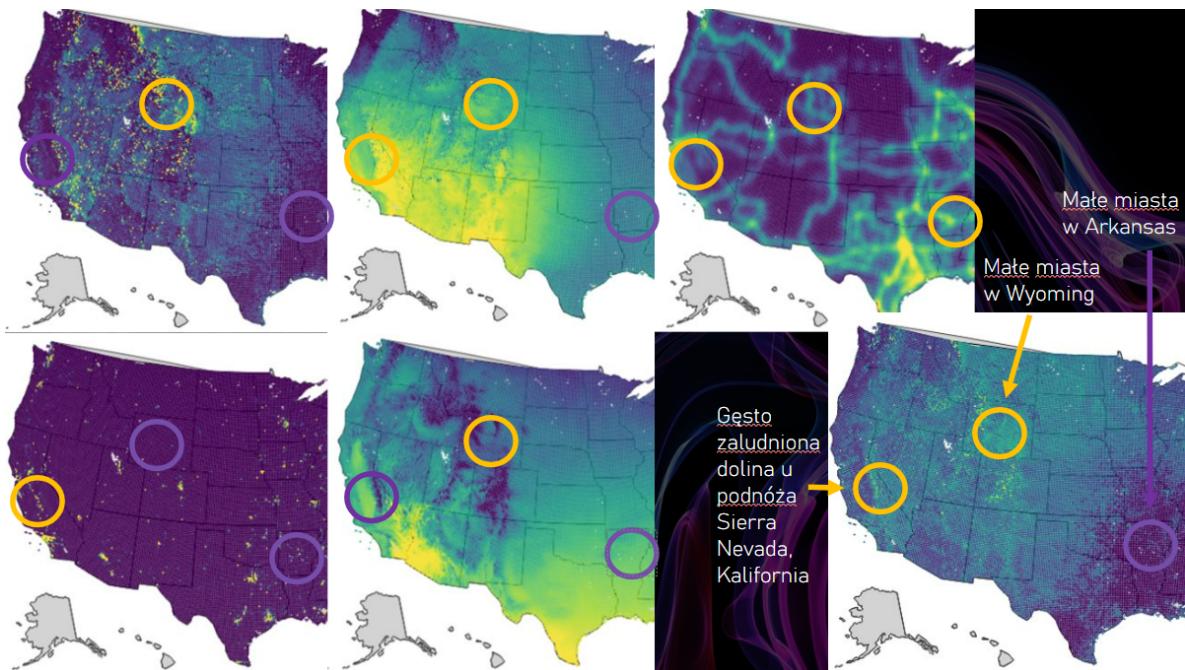
Została zaprojektowana aplikacja w wersji demo, która została skonstruowana w taki sposób aby wyświetlać wszystkie zgromadzone dane, wyświetlać wyniki modelu, oraz dokonywać lokalnej optymalizacji określonej liczby rozmieszczonej rozproszonych Data Center za pomocą algorytmu ewolucyjnego.



Rysunek 7: Przykładowe miejsca i powiązania z atrybutami



Rysunek 8: Przykładowe miejsca i powiązania z atrybutami



Rysunek 9: Przykładowe miejsca i powiązania z atrybutami