

Symulacja migracji ludności na terenie Stanów Zjednoczonych

Projekt na metody stochastyczne

Filip Ręka
Daniel Kuc
Piotr Rzeźnik

A
G
H

24 czerwca 2023

Wstęp

Migracja jest powszechna w dzisiejszym świecie. Ludzie migrują ze względu na pracę, szkołę, studia, rodzinę, a także przez wiele innych czynników. Migrację można podzielić ze względu na kilka aspektów, ale w tym przypadku najważniejszy jest podział na migrację wewnętrzną i zewnętrzną. Ta praca będzie opierała się na migracji wewnętrznej, czyli w obrębie granic jednego państwa. Stany Zjednoczone zostały wybrane ze względu na bardzo dużą mobilność ich ludności, czyli skłonność do zmiany miejsca zamieszkania [1].

Przegląd literatury

Temat migracji pojawia się w literaturze od wielu lat i jest przedmiotem badań w wielu obszarach naukowych. Michael Greenwood w swoim artykule rozważał trendy, kierunki i powody wewnętrznej migracji w Stanach Zjednoczonych w latach 1965-1983 [2]. Bardzo często badany jest wpływ zmiany klimatu i zagrożeń z niego płynących na migrację ludności [3]. Brany jest pod uwagę wzrost poziomu wody w morzach, powodzie, a także susze [4, 5, 6]. Często migracja jest zestawiana z rozprzestrzenianiem się chorób. Były przeprowadzone badania na temat wpływu migracji na rozprzestrzenianie się np. malarii lub koronawirusa [7, 8]. Do modelowania samej migracji wykorzystywane od dawna był wykorzystywany model grawitacyjny, który przewiduje stopień migracji pomiędzy dwoma miejscami i opiera się na prawie powszechnego ciężenia. Model ten jest nadal popularny w ostatnich latach [9]. Rozszerzeniem modelu grawitacyjnego jest model radiacyjny. Także bazuje on na zjawisku fizycznym, tym razem na rozchodzeniu się fal w próżni [10]. Może on zostać zgeneralizowany także na inne cechy dwóch miejsc np. ich poziom zurbanizowania [11]. W badaniach nad migracją ludności w sieci miejsc używano już łańcuchów Markowa do symulowania przepływu ludności między miastami [12]. Do przewidywania kierunków migracji ludności były użyte także metody uczenia maszynowego. W artykule Caleb Robinson i Bistra Dilkina zostały porównane ze sobą modele grawitacyjne, radiacyjne oraz modele uczenia maszynowego XGBoost i sieć neuronowa (MLP) [13].

Dane użyte przy realizacji projektu

Cechy, które zdecydowaliśmy się użyć to:

1. Procent ludzi w danym hrabstwie z wykształceniem wyższym
2. Zmiana indeksu HPI (*House price index*), mówiący o zmianie cen mieszkań
3. Średni przychód w hrabstwie
4. Bezrobocie w hrabstwie
5. Stan zdrowia
6. Stan zdrowia psychicznego
7. Ilość przestępczości

8. Wartość dotycząca problemów ze znalezieniem zamieszkania
9. Indeks nierówności społecznych
10. Wartość opisująca dostępność miejsc do ćwiczeń

Dane pochodzą z instytucji rządowych, przez co mamy pewność że są wiarygodne. Opisują stan rzeczy na rok 2016.

Aby móc wylosować cechy konkretnej osoby, należy najpierw znaleźć odpowiednią dystrybucję do każdej z cech. Nie możemy *a-priori* założyć że wszystkie dane pochodzą na przykład z rozkładu normalnego, gdzie średnia to wartość cechy danego hrabstwa.

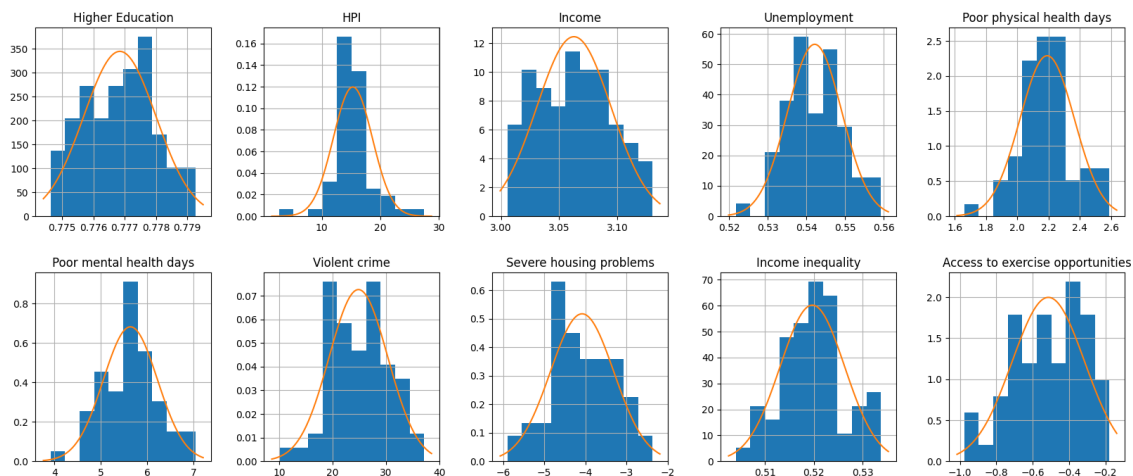
Udało nam się odszukać odpowiednie dystrybucje, które reprezentują dane na poziomie całych stanów zjednoczonych, co zostało pokazane na grafice poniżej.



Aby umożliwić losowanie z danymi rozkładami zastosowaliśmy transformację Boxa-Coxa. Jest to operacja która powoduje przetworzenie danych w taki sposób aby ich rozkład jak najbardziej był zbliżony do rozkładu normalnego. Transformacja jest wyrażona wzorem:

$$y(\lambda) = \begin{cases} (y^\lambda - 1)/\lambda, & \text{if } \lambda \neq 0 \\ \ln(y), & \text{if } \lambda = 0 \end{cases} \quad (1)$$

Na na poniższym obrazku znajdują się rozkłady po transformacji Boxa-Coxa:



Widać, że wszystkie rozkłady teraz są zbliżone do rozkładu normalnego. Na podstawie tych rozkładów mając średnią i odchylenie standardowe z klastrów na które podzieliliśmy stany możemy losować populację każdego hrabstwa. Transformacja jest odwracalna więc po wylosowaniu możemy powrócić do bazowego zakresu wartości cech.

Kroki przy realizacji projektu

Wybór cech wpływających na migrację w danym hrabstwie

Mając dane dotyczące ilości osób które migrują z danego hrabstwa do innych hrabstw, jesteśmy w stanie dowiedzieć się, które cechy z jaką siłą wpływają na migrację ludności. Oczywiście jest to, że w każdym hrabstwie będą to inne cechy, dlatego prezentujemy takie podejście do rozwiązania tego problemu.

Każde hrabstwo jest przedstawione przy pomocy wektora cech o stałym rozmiarze N . Mając dane dotyczące ilości osób, które z rozważanego hrabstwa wyemigrowały, jesteśmy w stanie policzyć siłę "przyciągania" w następujący sposób:

- h_A - macierz cech hrabstwa z którego emigrację rozważamy
- h_1, h_2, \dots, h_k - macierze cech hrabstw, do których mieszkańcy hrabstwa A emigrowali

Dla każdego hrabstwa od 1 do k posiadamy dane dotyczące ile osób z hrabstwa A wyemigrowało do każdego z pozostałych. Wpiszmy te dane do wektora d . Aby obliczyć prawdopodobieństwa stosunku migracji używamy funkcji softmax, która zamieni nasz wektor w wektor prawdopodobieństw.

$$t = \text{softmax}(d) \quad (2)$$

Następnie jesteśmy w stanie policzyć "wirtualne" hrabstwo, w następujący sposób:

$$h_{vA} = \sum_{i=1}^k t_i h_i \quad (3)$$

"Wirtualne" hrabstwo jest średnim hrabstwem, do którego przeciętna osoba z danego hrabstwa będzie chciała emigrować.

Jeżeli każde hrabstwo jest opisywane przez wektor o stałej długości, jesteśmy w stanie nanieść wszystkie hrabstwa na N -wymiarową przestrzeń. Osoba zamieszkująca dane hrabstwo jest opisywana przy pomocy takiej samej liczby cech co wektor dla hrabstwa. Osoby, które są bardziej skłonne do migracji, po naniesieniu ich wektora cech na przestrzeń, będą znajdowały się dalej od rodzinnego hrabstwa, odległość ta determinuje skłonność do migracji. Podejście to nie bierze pod uwagę czynników, które mogą wpłynąć na miejsce docelowe. Przykładowo osoba znajdująca się w hrabstwie o niskim PKB, będzie bardziej skłonna do emigracji do tych hrabstw, które również mają niskie PKB, co jest raczej nieintuicyjne, ponieważ będzie znajdowała się w przestrzeni bliżej nich.

Musieliśmy znaleźć możliwość zmienienia wartości cech danej osoby na podstawie danych z migracji z danego hrabstwa. Pierwszym pomysłem było wykorzystanie sieci neuronowej, która zostałaby nauczona aby zamieniać wartości danego hrabstwa na wartości hrabstwa "wirtualnego". W ten sposób, wiedząc z jakiego hrabstwa pochodzi,

moglibyśmy użyć modelu, a następnie nanieść na przestrzeń zamienione wartości danej osoby.

Aby lepiej zachować lokalne warunki, nie moglibyśmy nauczyć jednej sieci na całe Stany Zjednoczone (która może pominąć lokalne potrzeby osób emigrujących), oraz nie moglibyśmy nauczyć jednej sieci dla danego hrabstwa (ponieważ mamy tylko jedną próbkę do nauki). Zamiast tego, pogrupowalibyśmy wektory hrabstw przy pomocy algorytmu K-means i wytrnowalibyśmy jedną sieć neuronową na każdy klaster. W ten sposób zostałyby zachowany balans pomiędzy lokalnością sieci oraz ilością danych do nauki.

Niestety w implementacja sieci i wykorzystanie jej na prawdziwych danych zweryfikowało ten pomysł. Sieć nie dawała dobrych rezultatów i osoby z danego hrabstwa przesuwały się w przestrzeni cech w to samo miejsce, co uniemożliwiało przeprowadzenie prawdziwej symulacji migracji.

Drugim pomysłem było wyliczenie wag cech istotnych dla osób migrujących z danego hrabstwa. Biorąc pod uwagę cechy danego hrabstwa i jego wirtualnego hrabstwa obliczaliśmy istotność poszczególnych cech. Wektor istotności został policzony w następujący sposób:

$$h_{vA} = 1 + (h_{vA} - h_A) \quad (4)$$

Następnie cechy osób z tego hrabstwa zostały przemnożone przez powyższy wektor wag przesuwając osoby w przestrzeni cech bliżej miejsc bardziej im odpowiadających. To podejście sprawdziło się i zostało wykorzystane w projekcie.

Przebieg symulacji

1. Tworzymy strukturę KD-Tree, w którym umieszczamy hrabstwa na podstawie wartości ich cech.
2. Losujemy ustalonej liczby przedstawicieli z każdego hrabstwa.
3. Cechy każdej wylosowanej osoby przekształcamy wektorem istotności cech.
4. Dla każdego mieszkańca obliczymy jego “odległość” w przestrzeni cech od hrabstwa rodzinnego i na podstawie tej odległości decydujemy czy będzie on migrował czy nie.
5. Losujemy bez zwracania liczbę emigrantów na podstawie łącznej wartości migracji z danego hrabstwa.
6. Dla każdej osoby wyciągamy z KD-Tree ustaloną liczbę najbliższych hrabstw oraz odległości do nich i na podstawie tych odległości tworzymy wektor prawdopodobieństwa migracji do każdego hrabstwa. Następnie z wyliczonym prawdopodobieństwem losujemy jedno hrabstwo do którego reprezentant wyemigruje.
7. Przez ustaloną liczbę kroków symulacji wykonujemy polecenia zaczynając od numeru 3.

Ilustracja mechanizmu projektu

$$\mathbf{A} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}, \mathbf{B} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix}, \mathbf{C} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix}, \mathbf{D} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix}, \mathbf{E} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}, \mathbf{F} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{bmatrix} \quad (5)$$

A, B, C, D, E oraz F są przykładowymi hrabstwami, mające odpowiadające im cechy zapisane w wektorach po prawej stronie znaku równa się.

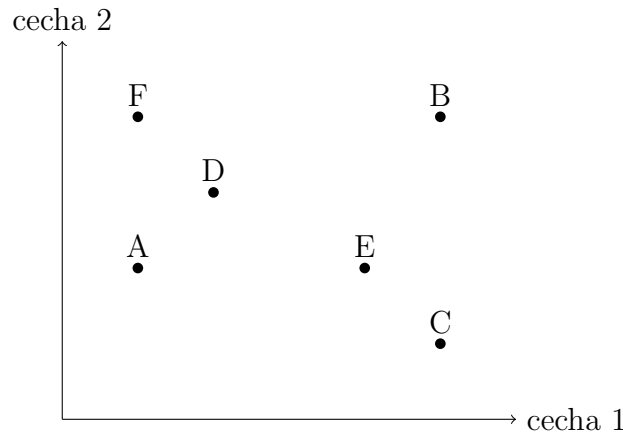
Na potrzeby prezentacji, założymy, że osoby z hrabstwa A migrowały w następujących ilościach: x_1, x_2, x_3 kolejno do hrabstw B, C oraz D.

Policzmy “wirtualne” hrabstwo V_A dla A:

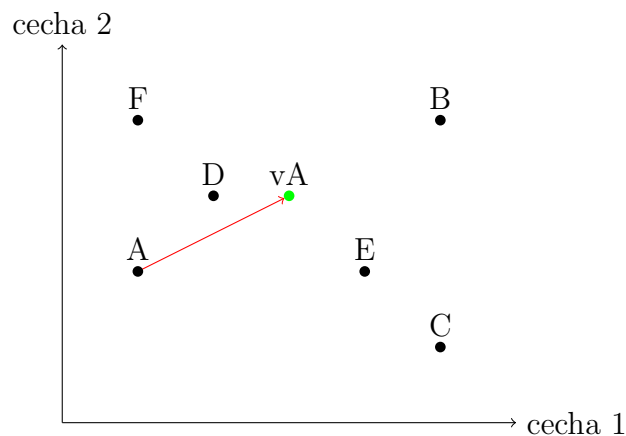
$$\begin{aligned} t &= \text{softmax}([x_1, x_2, x_3]) \\ V_A &= t_1 B + t_2 C + t_3 D \\ V_A &= \begin{bmatrix} v_{a1} \\ v_{a2} \\ \vdots \\ v_{aN} \end{bmatrix} \end{aligned} \quad (6)$$

Analogiczne operacje zostaną wykonane aby policzyć “wirtualne” hrabstwo dla każdego z naszych hrabstw.

Nanieśmy teraz nasze hrabstwa na przestrzeń. Dla uproszczenia wizualizacji pokażemy tylko ją dla dwóch wymiarów.



Nanieśmy także przykładowe wirtualne hrabstwo na daną przestrzeń.



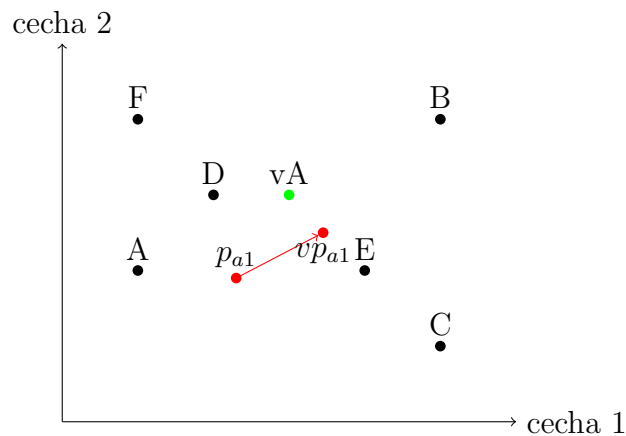
To oznacza że osoby migrujące z hrabstwa A wybierają hrabstwa o większej wartości cechy 1 oraz o większej wartości cechy 2. Wirtualne hrabstwo A znajduje się blisko hrabstwa D oraz hrabstwa E. Wektor wag istotności cech obliczamy tak jak to zostało przedstawione w rozdziale Kroki realizacji projektu.

Następnie losujemy populację osób z każdego hrabstwa, proporcjonalną do wielkości hrabstwa.

Przyjrzyjmy się osobie p_{a1} , która została wylosowana z hrabstwa A. Z hrabstwem A jest skojarzony odpowiedni wielowymiarowy wspólny rozkład prawdopodobieństwa f_A , który jest w odpowiedni sposób sparametryzowany, co zostało opisane w sekcji poświęconej temu tematowi.

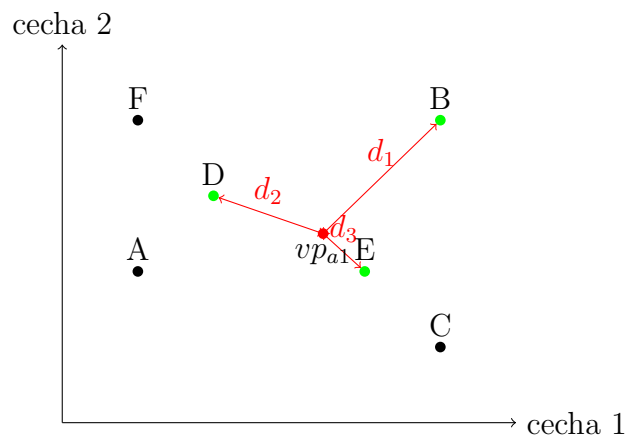
$$h_{a1} \sim f_A \quad (7)$$

Następnie wektor cech osoby zostaje przekształcony za pomocą wektora wag istotności cech osób w danym hrabstwie.



Na powyższym rysunku przedstawiono wszystkie hrabstwa, wylosowaną osobę, a jej transformację przedstawiono przy pomocy strzałki.

Następnie znajdujemy k najbliższych sąsiadów (hrabstw) dla punktu vp_{a1} i obliczamy do nich odległości. Oznaczmy je d_1, d_2, \dots, d_k .



Losujemy do którego z hrabstw osoba h_{a1} wyemigruje z prawdopodobieństwami $\text{softmax}([d_1, d_2, \dots, d_k])$.

Aktualizujemy wszystkim osobom domowe hrabstwo oraz cechy i wykonujemy kolejny obieg symulacji. Czynności są powtarzane przez ustaloną liczbę kroków.

Wyniki

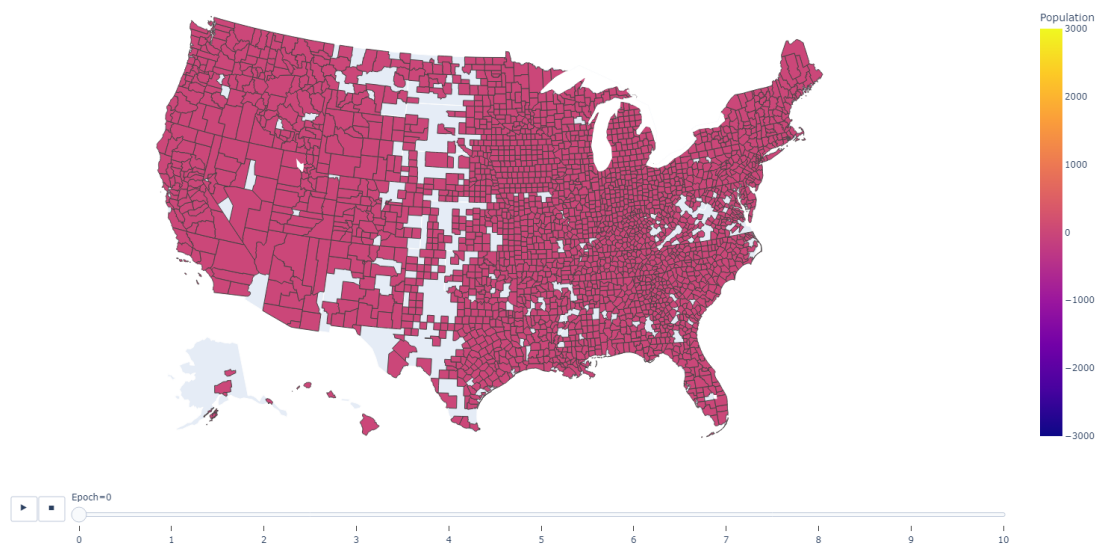
Przeprowadziliśmy symulację dla różnych liczb najbliższych sąsiadów, spośród których losowane byłoby hrabstwo do którego osoba z populacji migruje. Sprawdzone przez nas liczby najbliższych sąsiadów to 10, 100, 500 i 1000. W każdym z wariantów przeprowadziliśmy 10 iteracji symulacji. Wyniki są zróżnicowane i pokazują różne aspekty migracji w Stanach Zjednoczonych.

Epoch	k=10	k=100	k=500	k=1000
1	115.0	92.9	72.4	69.2
5	481.5	405.0	319.6	298.2
10	817.2	699.9	557.7	513.4

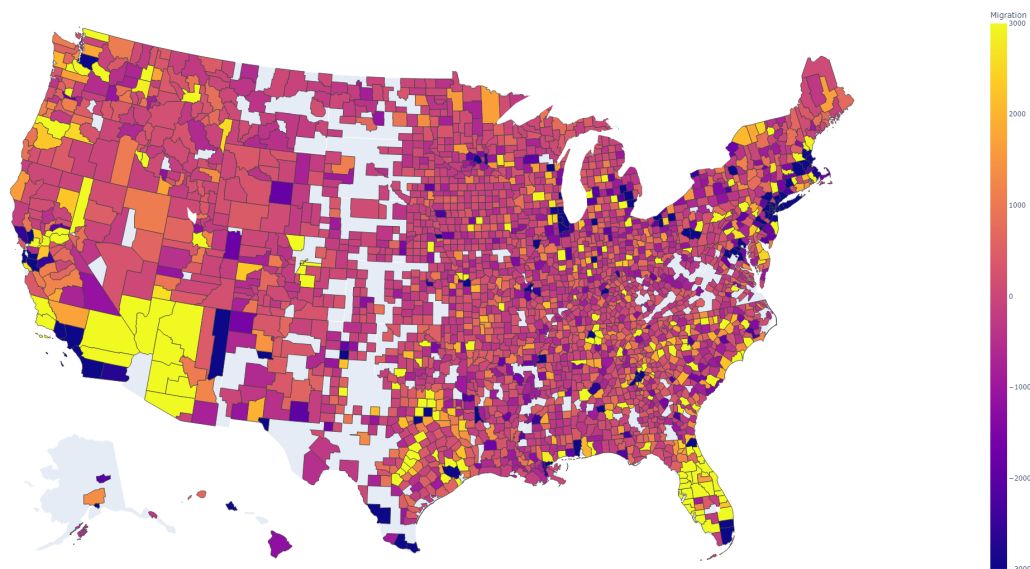
Tabela 1: Błąd średniokwadratowy dla wybranych epok i różnej liczby najbliższych sąsiadów.

Według powyższej tabeli błąd zmierzony pomiędzy wynikiem symulacji, a prawdziwą wartością migracji rośnie z kolejnymi iteracjami symulacji. Błąd średniokwadratowy jest mniejszy dla losowania spośród większej liczby najbliższych sąsiadów, aczkolwiek skutkuje to zwiększeniem czasu obliczeń symulacji.

Na poniższym obrazku widzimy stan początkowy. Na szaro są hrabstwa z których nie posiadaliśmy danych.

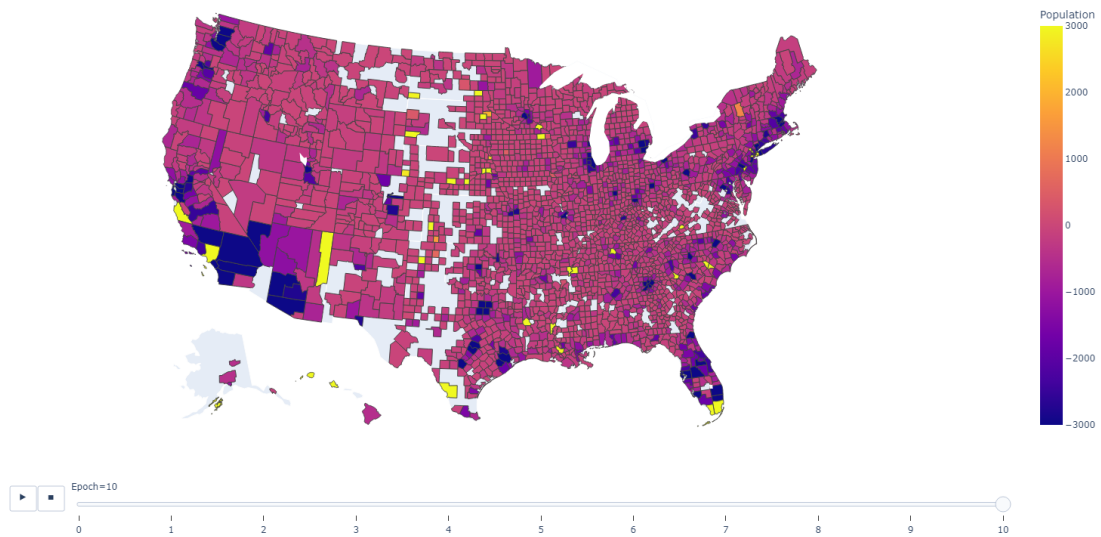


Za punkt odniesienia bierzemy mapę rzeczywistych migracji:



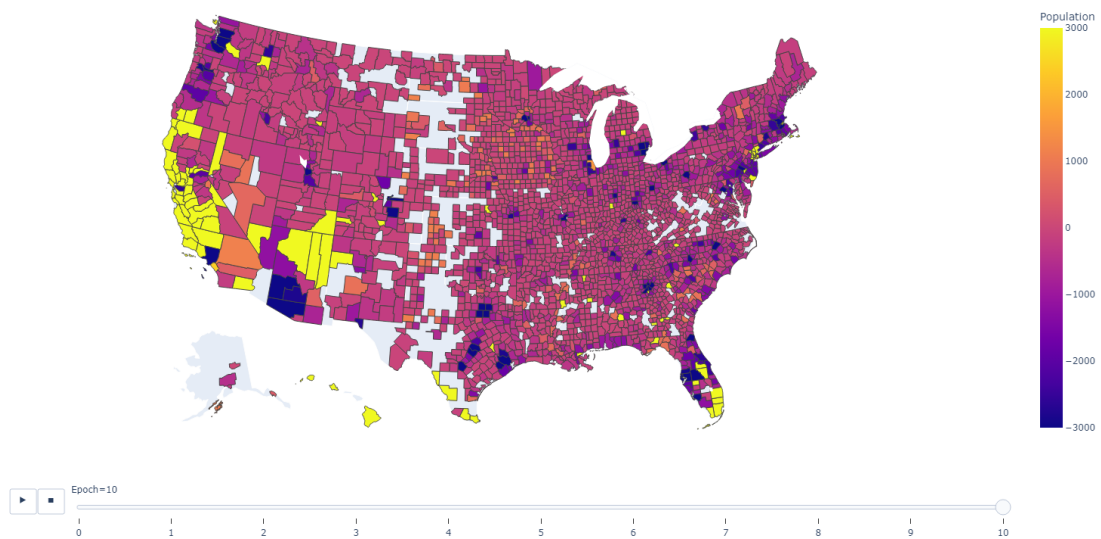
Widać na niej odpływ ludności z dużych miast do okolicznych hrabstw. Widać to zwłaszcza na Florydzie, w Arizonie i w Kalifornii. Widać także przemieszczanie się ludności pomiędzy hrabstwami w środkowej części Stanów oraz przyływ ludzi do małych miast.

Mapa dla 10 najbliższych sąsiadów podczas wybierania kierunku migracji:



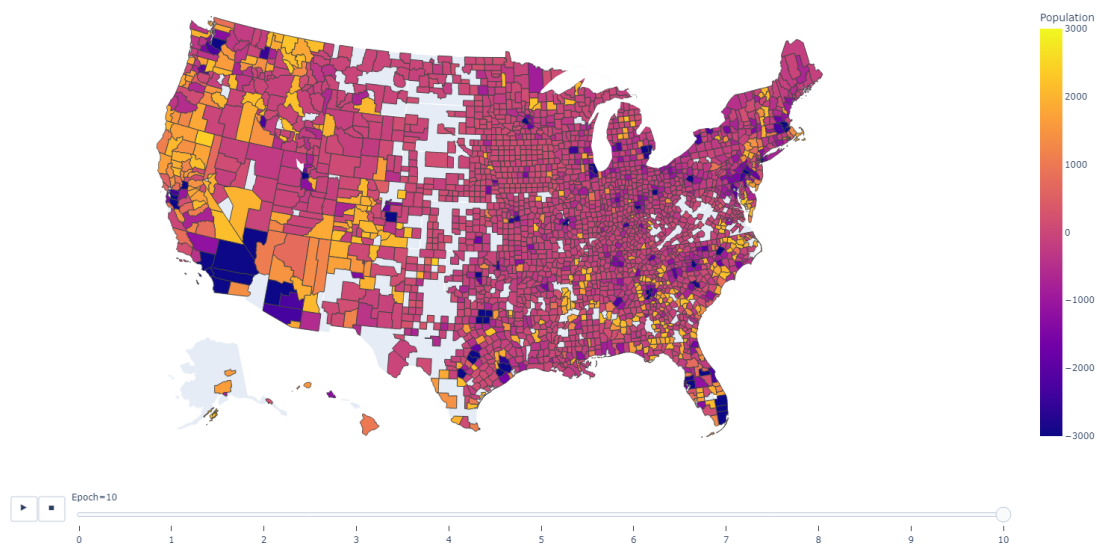
Dla 10 najbliższych sąsiadów skupiamy się głównie na ogólnie najlepszych hrabstwach, do których chciałyby migrować osoby z całych stanów. Dlatego na mapie przedstawiającej stan po migracjach widać pojedyncze mocno żółte hrabstwa które zerbały bardzo dużą liczbę migrantów. Wynika z tego że zbyt duża lokalność zamyka możliwości do migracji w pewne mniej oczywiste miejsca przez co źle oddaje prawdziwą dynamikę migracji.

Mapa dla 100 najbliższych sąsiadów podczas wybierania kierunku migracji:



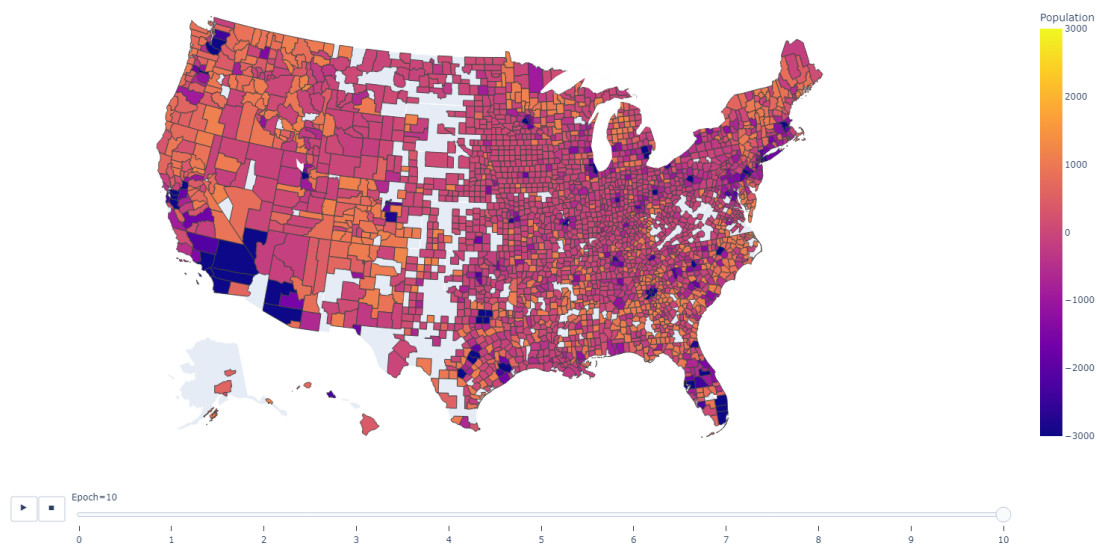
Powyższa mapa lepiej obrazuje rzeczywistą migrację. Widzimy, że niektóre duże ośrodki notują odpływ społeczeństwa, a okoliczne hrabstwa notują przyływ ludności. Na zachodnim wybrzeżu przybywa ludzi także jak w rzeczywistości.

Mapa dla 500 najbliższych sąsiadów podczas wybierania kierunku migracji:



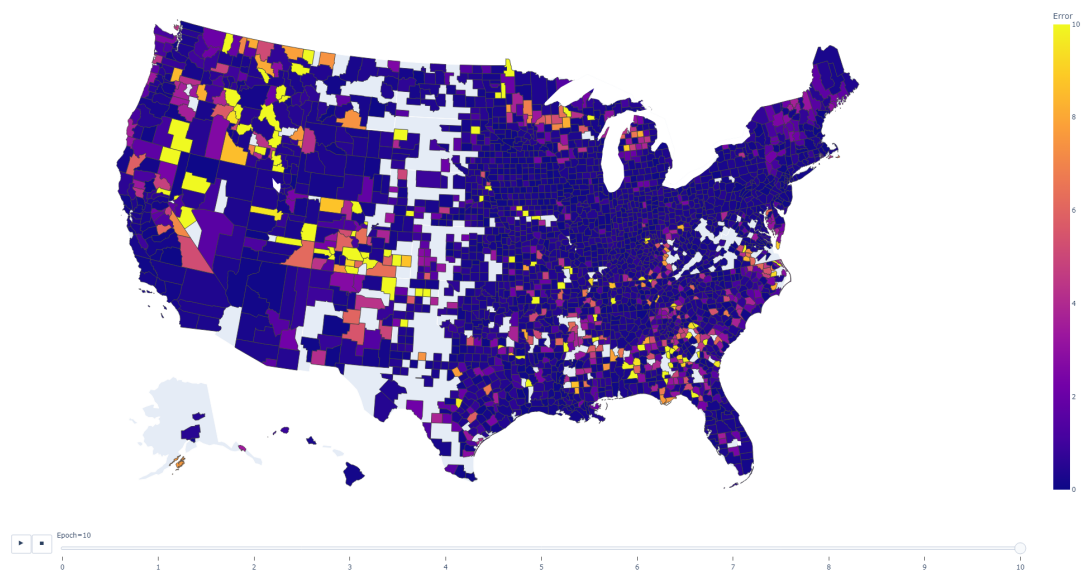
W przypadku 500 sąsiadów możemy zaobserwować odmienne schematy niż w dwóch poprzednich przypadkach. Część obszarów pokrywa się z prawdziwymi danymi. Widać, że nie ma tutaj znacznego nasycania się pewnych konkretnych miejsc a osoby rozpraszają się po wielu hrabstwach.

Mapa dla 1000 najbliższych sąsiadów podczas wybierania kierunku migracji:



Na tej mapie widnieje większa tendencja chęci wyjazdu z pewnych konkretnych hrabstw. Nie ma na tej mapie obszarów z nasyconym żółtym kolorem. Ludzie opuszczali pewne miejsca zaznaczone na niebiesko i rozpraszali się po całych stanach. W mniej licznych hrabstwach w centralnej części ameryki nastąpiło przemieszanie ludności.

Na poniższej mapie znajduje się procentowa różnica pomiędzy przewidywaną wartością populacji w każdym hrabstwie, a prawdziwą wartością populacji dla najlepszego modelu względem błędu średniokwadratowego - k1000:



Widać, że w większości hrabstw populacje osiągnęły poziom zbliżony do rzeczywistego. Występują pewne hrabstwa, w których zmiana populacji wykazała się inną tendencją niż miało to miejsce w rzeczywistości.

Wnioski

Migracja jest procesem bardzo złożonym. Udało się nam dostać wyniki częściowo zbieżne z prawdziwymi danymi, co jest dużym sukcesem. Można na nich odnaleźć wzorce i mechanizmy widoczne w prawdziwym procesie migracji. Oczywiście są także rozbieżności pomiędzy rzeczywistością, a wynikami. Wpływ na to ma wiele czynników, między innymi ilość i dobór cech. Wybraliśmy jedynie 10 zmiennych, które mogą zbyt słabo opisywać powody społeczeństwa do migracji. Możliwe, że nasz wybór cech spowodował, że w symulacji pewne hrabstwa były wybierane częściej/rzadziej przez osobników, a w rzeczywistości są także inne wskaźniki, przez które ludzie chcieliby bądź nie chcieliby migrować do danego hrabstwa.

Podsumowując, projekt pozwolił zrozumieć pewne czynniki stojące za migracją oraz Jesteśmy w stanie, w pewnym stopniu poprawnie, zasymulować migrację zachowując wzorce i tendencje migracyjne. W przyszłości projekt możnaby było rozszerzyć o szerszy zestaw cech, opisujących obszary i populację. Można też zastosować model symulacyjny na inne państwa, a nawet cały świat.

Bibliografia

- [1] Wikipedia contributors. *Migracja ludności — Wikipedia, wolna encyklopedia*. [Online;] 2023. URL: https://pl.wikipedia.org/wiki/Migracja_ludno%C5%9Bci.
- [2] Michael J. Greenwood. “Human migration: theory, models, and empirical studies.” In: *Journal of regional science* 25 4 (1985), pp. 521–44.
- [3] Rebecca Parrish et al. “A Critical Analysis of the Drivers of Human Migration Patterns in the Presence of Climate Change: A New Conceptual Model”. In: *International Journal of Environmental Research and Public Health* 17.17 (2020). ISSN: 1660-4601. DOI: 10.3390/ijerph17176036. URL: <https://www.mdpi.com/1660-4601/17/17/6036>.
- [4] Sabine Perch-Nielsen. “Understanding the effect of climate change on human migration”. In: (2004). URL: <https://www.research-collection.ethz.ch/>.
- [5] Pietro De Lellis, Manuel Ruiz Marín, and Maurizio Porfiri. “Modeling Human Migration Under Environmental Change: A Case Study of the Effect of Sea Level Rise in Bangladesh”. In: *Risk Analysis* (2021).
- [6] Sibren Isaacman, V. F.-M., and E. F.-M. “Modeling Human Migration Patterns During Drought Conditions in La Guajira, Colombia”. In: *Conference on Computing and Sustainable Societies (COMPASS '18) Proceedings* (2018).
- [7] N.N.A.B.M.E.K.H. Rejeki. “A Time Series Analysis: Weather Factors, Human Migration and Malaria Cases in Endemic Area of Purworejo, Indonesia, 2005-2014”. In: *Iranian Journal of Public Health* (2018).
- [8] T. C. F. Y. L. Z. Z. H. Zhan C. “Modeling and prediction of the 2019 coronavirus disease spreading in China incorporating human migration data.” In: *PLOS ONE* (2020).
- [9] Jacques Poot Omoniyi Alimi Michael P. Cameron David C. Maré. “The Gravity Model of Migration: The Successful Comeback of an Ageing Superstar in Regional Science”. In: *Investigaciones Regionales - Journal of Regional Research IZA DP No. 10329* (2016).
- [10] Maritan Amos Simini Filippo González Marta C. and Barabási Albert-László. “A Universal Model for Mobility and Migration Patterns”. In: *Nature* 484.7392 (2012), pp. 96–100. ISSN: 1476-4687. DOI: 10.1038/nature10856. URL: <https://doi.org/10.1038/nature10856>.
- [11] Legara Erika Fille Alis Christian and Monterola Christopher. “Generalized Radiation Model for Human Migration”. In: *Scientific Reports* 11.1 (2021), p. 22707. ISSN: 2045-2322. DOI: 10.1038/s41598-021-02109-1. URL: <https://doi.org/10.1038/s41598-021-02109-1>.
- [12] J. Pan and A. Nagurney. “Using Markov chains to model human migration in a network equilibrium framework”. In: *Mathematical and Computer Modelling* 19.11 (1994), pp. 31–39. ISSN: 0895-7177. DOI: [https://doi.org/10.1016/0895-7177\(94\)90014-0](https://doi.org/10.1016/0895-7177(94)90014-0). URL: <https://www.sciencedirect.com/science/article/pii/0895717794900140>.
- [13] Caleb Robinson and Bistra Dilkina. *A Machine Learning Approach to Modeling Human Migration*. 2017. arXiv: 1711.05462 [cs.SI].