

Dane geo-lokalizowane stanowią większość danych generowanych współcześnie. Ich stały napływ do wykorzystania w badaniach empirycznych tworzy bodźce do rozwoju metodologii ich analizy, jak również do tworzenia teorii ekonomicznych zasadzających się na nowych spostrzeżeniach. Niniejszy projekt jest odpowiedzią na te potrzeby badawcze.

Z perspektywy metodologicznej, analiza punktowych danych geo-lokalizowanych wymaga szczególnych metod ilościowych uwzględniających sąsiedztwo i położenie, jak również skalowalnych do *big data*. Metody klasycznej ekonometrii przestrzennej, uwzględniające autokorelację przestrzenną (podobieństwo do sąsiadów) projektowane dla regionów nie mogą w prosty sposób być implementowane dla danych punktowych, z grid czy ze zdjęć satelitarnych ze względu na ograniczenie skalowalności oraz wymaganą stałość macierzy wag przestrzennych. Coraz powszechniej próbuje się wykorzystywać metody uczenia maszynowego, choć wciąż są to zastosowania a-przestrzenne nieuwzględniające autokorelacji przestrzennej. Wyzwaniem modelowania są dane punktowe, które wykazują bardzo zróżnicowaną gęstość przestrzenną, zmienną lokalizację i duży wolumen.

Z perspektywy teoretycznej, zróżnicowana relatywna gęstość przestrzenna zjawisk, głównie lokalizacji firm, jest słabo zbadana i wciąż nie stanowi czynnika tłumaczącego procesy gospodarcze. W tym projekcie stawiana jest **hipoteza, że korzyści gęstości mogą skutecznie tłumaczyć decyzje lokalizacyjne firm z różnych sektorów, głównie w układzie rdzeń-peryferia**. Konieczna jest empiryczna weryfikacja tej hipotezy, co może rzucić nowe światło na mechanizmy procesów gospodarczych i decyzyjnych przedsiębiorstw w skali mikro, a także wypracować dobry sposób detekcji potencjalnej korzystnej lokalizacji dla nowych firm z wybranego sektora.

Niniejszy projekt zakłada kompleksowe badania nad problemem zmiennej gęstości lokalizacji firm w ramach trzech zadań: 1) prac metodologicznych, 2) prac teoretycznych i 3) prac empirycznych. Zakładany wkład projektu w rozwój nauki:

- **w zakresie metodologii metod ilościowych:** stworzenie koncepcji przestrzennego modelu przełącznikowego, pozwalającego na modelowanie endogeniczności przestrzennej wynikającej z przynależności punktów do klastrów różnej gęstości, poprawę skalowalności metod ekonometrycznych i lepsze dopasowanie modelu w sytuacji heterogeniczności przestrzennej. Jest to podejście pośrednie pomiędzy modelem globalnym z jednym współczynnikiem i modelem ze współczynnikami indywidualnymi dla każdej obserwacji (GWR, *Geographically Weighted Regression*). Model przełącznikowy jest podejściem zbieżnym z często wykorzystywanym partycjonowaniem dużych zbiorów danych w celu paralelizacji obliczeń i pozwala na znaczące podniesienie skalowalności obliczeń przy zachowaniu wykonalności obliczeń na sprzęcie biurowym i w typowych pakietach do analiz przestrzennych.
- **w zakresie metodologii metod ilościowych:** wypracowanie metodologii integrującej przestrzenne uczenie maszynowe nadzorowane i nienadzorowane oraz ekonometrię przestrzenną w celu auto-klasyfikacji nowych punktów do klastrów i predykcji potencjału lokalizacyjnego obszaru, poprawy jakości prognoz i samo-uczenia się modelu
- **zakresie teorii ekonomii:** rozwinięcie teorii korzyści gęstości, dotychczas stosowanej głównie w logistyce, w powiązaniu z teoriami wyboru lokalizacji przez firmy oraz z korzyściami aglomeracji, urbanizacji, koncentracji
- **w zakresie badań empirycznych:** wypracowanie pilotażowego modelu detekcji potencjalnej lokalizacji firm z wybranych sektorów oraz prognozowania potencjału lokalizacyjnego danego terytorium

Wyniki projektu mogą znaleźć zastosowanie w badaniach nad lokalizacją firm, organizacją miasta i przestrzeni, spójnością terytorialną, problemami aglomeracji, zatłoczenia, efektów zewnętrznych, w tym środowiskowych, preferencjami osadniczymi, w geo-marketingu, projektowaniu obszarów zasięgu (tzw. *catchment areas*) czy modelowaniu mobilności rejestrowanej w telefonii komórkowej.