

Czy da się przewidzieć ruch turystyczny w polskich powiatach?

Analiza i modelowanie liczby udzielonych noclegów na podstawie danych BDL GUS

Uniwersytet Pomorski w Słupsku

Autorzy:

Uladzislau Vilchuk

vladvilchuk1@gmail.com

Andrii Shevchenko

andrzejewczenko27.08.2007@gmail.com

Nazar Haievskyj

49203@office.upsl.edu.pl

Daniil Zubkov

zubkovdaniil86@gmail.com

Streszczenie

Ruch turystyczny stanowi istotny element rozwoju społeczno-gospodarczego regionów, wpływając na poziom zatrudnienia, rozwój infrastruktury oraz dochody jednostek samorządu terytorialnego. Jednym z podstawowych mierników intensywności tego zjawiska jest liczba udzielonych noclegów, która odzwierciedla zarówno skalę przyjazdów turystów, jak i długość ich pobytu.

Celem niniejszej pracy jest analiza zróżnicowania przestrzennego i czasowego ruchu turystycznego w polskich powiatach oraz ocena możliwości prognozowania liczby udzielonych noclegów na podstawie danych statystycznych Banku Danych Lokalnych Głównego Urzędu Statystycznego (BDL GUS). Analiza obejmuje lata 2013–2024 i została przeprowadzona z wykorzystaniem metod analizy danych oraz modelowania predykcyjnego.

W ramach pracy przeanalizowano zmiany liczby noclegów w czasie, zidentyfikowano powiaty o najwyższym natężeniu ruchu turystycznego oraz oceniono wpływ wybranych czynników społeczno-gospodarczych i infrastrukturalnych na popyt turystyczny. Następnie przeprowadzono porównanie skuteczności czterech modeli regresyjnych, spośród których jako model docelowy wybrano algorytm Gradient Boosting Regressor, charakteryzujący się najwyższą trafnością prognoz w ujęciu lokalnym.

Uzyskane wyniki potwierdzają wysoką użyteczność danych BDL GUS w analizach regionalnych oraz wskazują, że zastosowanie zaawansowanych metod uczenia maszynowego

pozwala na skuteczne prognozowanie ruchu turystycznego na poziomie powiatów, przy jednoczesnym uwzględnieniu zróżnicowania przestrzennego badanego zjawiska.

Słowa kluczowe: turystyka, ruch turystyczny, noclegi, powiaty, BDL GUS, prognozowanie, Gradient Boosting

1. Wstęp

Turystyka odgrywa istotną rolę w funkcjonowaniu współczesnych gospodarek, stanowiąc jedno z ważniejszych źródeł dochodów dla wielu regionów oraz wpływając na rozwój infrastruktury transportowej, noclegowej i rynku pracy. Intensywność ruchu turystycznego oddziałuje zarówno na obszary miejskie, pełniące funkcje administracyjne i biznesowe, jak i na regiony o charakterze wypoczynkowym, w szczególności obszary górskie i nadmorskie.

W Polsce ruch turystyczny charakteryzuje się znacznym zróżnicowaniem przestrzennym. Analiza tego zjawiska na poziomie lokalnym, w tym w ujęciu powiatowym, umożliwia lepsze zrozumienie mechanizmów kształtujących popyt turystyczny oraz stanowi istotne wsparcie dla procesów planowania rozwoju regionalnego. W badaniach nad turystyką powszechnie wykorzystuje się liczbę udzielonych noclegów jako podstawowy wskaźnik intensywności ruchu turystycznego, gdyż łączy on informacje o liczbie przyjazdów oraz długości pobytu turystów.

Dynamika ruchu turystycznego jest jednak silnie uzależniona od czynników zewnętrznych, takich jak koniunktura gospodarcza, zmiany regulacyjne czy zdarzenia losowe o charakterze globalnym. Szczególnym przykładem takiego czynnika była pandemia COVID-19, która doprowadziła do gwałtownego spadku liczby noclegów w latach 2020–2021 oraz istotnie zaburzyła dotychczasowe trendy rozwoju turystyki w Europie i w Polsce, co potwierdzają raporty międzynarodowych organizacji, w tym UNWTO.

W ostatnich latach coraz większe znaczenie w analizach ekonomicznych i regionalnych zyskują metody uczenia maszynowego, umożliwiające modelowanie złożonych i nieliniowych zależności pomiędzy zmiennymi. W kontekście badań nad ruchem turystycznym podejście to pozwala nie tylko na analizę historycznych trendów, lecz również na skuteczniejsze prognozowanie przyszłych zmian na poziomie lokalnym, w tym w ujęciu powiatowym, przy wykorzystaniu danych administracyjnych i statystyki publicznej.

2. Cel i zakres pracy

Celem głównym niniejszej pracy jest analiza zróżnicowania przestrzennego i czasowego liczby udzielonych noclegów w polskich powiatach w latach 2013–2024 oraz ocena możliwości skutecznego prognozowania ruchu turystycznego na podstawie danych Banku Danych Lokalnych GUS.

Cele szczegółowe obejmują:

- analizę zmian liczby udzielonych noclegów w Polsce w ujęciu czasowym,
- identyfikację powiatów o najwyższym natężeniu ruchu turystycznego,
- ocenę wpływu wybranych czynników społeczno-gospodarczych i infrastrukturalnych na liczbę noclegów,
- budowę i validację modelu predykcyjnego służącego do prognozowania liczby udzielonych noclegów na poziomie powiatów.

Zakres przestrzenny analizy obejmuje wszystkie powiaty w Polsce, natomiast zakres czasowy został dostosowany do dostępności danych i obejmuje lata 2013–2024.

3. Przegląd literatury

Ruch turystyczny stanowi jeden z istotnych elementów rozwoju społeczno-gospodarczego regionów, wpływając zarówno na rynek pracy, jak i na infrastrukturę lokalną oraz dochody jednostek samorządu terytorialnego. W literaturze przedmiotu podkreśla się, że analiza dynamiki turystyki wymaga uwzględnienia wielu czynników o charakterze demograficznym, ekonomicznym i infrastrukturalnym (UNWTO, 2023).

W ostatnich latach coraz większą rolę w badaniach nad ruchem turystycznym odgrywają metody ilościowe oraz podejścia oparte na analizie dużych zbiorów danych statystycznych. Dane administracyjne, gromadzone przez instytucje publiczne, takie jak Główny Urząd Statystyczny, umożliwiają prowadzenie analiz przestrzennych na poziomie regionów i powiatów, co pozwala na identyfikację zróżnicowania natężenia ruchu turystycznego w skali kraju (GUS, 2024).

Jednym z najczęściej stosowanych mierników intensywności turystyki jest liczba udzielonych noclegów, która odzwierciedla zarówno skalę przyjazdów turystycznych, jak i długość pobytu odwiedzających. W literaturze wskazuje się, że wskaźnik ten charakteryzuje się wysoką wrażliwością na czynniki zewnętrzne, takie jak sytuacje kryzysowe, zmiany gospodarcze czy ograniczenia mobilności (Eurostat, 2022). Szczególnym przykładem takiego zjawiska była pandemia COVID-19, która spowodowała gwałtowny spadek liczby noclegów w większości krajów europejskich, w tym również w Polsce.

Coraz częściej w badaniach nad turystyką wykorzystuje się modele statystyczne i metody uczenia maszynowego do prognozowania przyszłych wartości ruchu turystycznego. Modele regresyjne, oparte na zmiennych demograficznych, gęstości zaludnienia, dostępności infrastruktury noclegowej czy liczbie obiektów turystycznych, pozwalają na estymację popytu turystycznego w ujęciu przestrzennym (Song & Li, 2008). Wskazuje się jednak, że skuteczność takich modeli zależy od jakości danych wejściowych oraz właściwego doboru cech objaśniających.

W kontekście analiz regionalnych istotne znaczenie ma również identyfikacja obszarów o skrajnie wysokim natężeniu ruchu turystycznego, takich jak duże aglomeracje miejskie czy regiony o ugruntowanej funkcji turystycznej. Jak zauważają liczne badania, obszary te często wykazują niestabilność trendów oraz większe odchylenia od modeli predykcyjnych, co wynika z ich podatności na czynniki losowe, wydarzenia masowe oraz zmiany w zachowaniach turystów (UNWTO, 2023).

Na tle przedstawionych badań niniejsza praca wpisuje się w nurt analiz ilościowych, wykorzystujących dane statystyczne BDL GUS do oceny dynamiki ruchu turystycznego na poziomie powiatów w Polsce. Szczególny nacisk położono na analizę liczby udzielonych noclegów w latach 2013–2024 oraz na próbę prognozowania ich wartości z wykorzystaniem modelu regresyjnego.

Na tle przedstawionych badań niniejsza praca wpisuje się w nurt analiz ilościowych wykorzystujących dane statystyczne do oceny dynamiki ruchu turystycznego na poziomie regionalnym. W literaturze przedmiotu podkreśla się rosnącą rolę modeli zespołowych, takich jak wzmacnianie gradientowe, w zadaniach predykcyjnych opartych na danych regionalnych, ze względu na ich zdolność do odwzorowywania złożonych zależności nieliniowych oraz wysoką trafność prognoz. [4,5]

4. Źródła danych i metodyka badania

Podstawowym źródłem danych wykorzystanych w pracy jest **Bank Danych Lokalnych Głównego Urzędu Statystycznego (BDL GUS)**. BDL jest oficjalnym systemem udostępniania danych statystycznych w Polsce i zawiera informacje z wielu obszarów, w tym z zakresu turystyki, demografii oraz środowiska.

W badaniu wykorzystano wskaźniki opisujące:

- liczbę udzielonych noclegów w turystycznych obiektach noclegowych,
- liczbę obiektów i miejsc noclegowych,
- liczbę turystów korzystających z noclegów,
- liczbę ludności,
- wybrane cechy środowiskowe (np. lesistość).

Dane zostały pobrane z BDL w postaci tabel statystycznych, a następnie poddane procesowi oczyszczania i agregacji. W kolejnym kroku utworzono scalony zbiór danych (tzw. master dataset), który posłużył do przeprowadzenia analiz eksploracyjnych oraz budowy modelu predykcyjnego. Analiza obejmowała metody wizualizacji danych, analizę korelacji oraz modelowanie regresyjne.

4.1. Metodyka badań

Analizę danych oraz modelowanie predykcyjne przeprowadzono w języku **Python**, w środowisku **Jupyter Notebook**. Do przetwarzania, agregacji oraz analizy danych statystycznych wykorzystano bibliotekę **Pandas**, natomiast proces budowy i uczenia modeli regresyjnych zrealizowano przy użyciu biblioteki **Scikit-learn**. Wizualizacje wyników, obejmujące m.in. wykresy trendów, porównania oraz zależności pomiędzy zmiennymi, wykonano z zastosowaniem bibliotek **Matplotlib** oraz **Seaborn**.

Algorytm Wzmocnienia Gradientowego został zaproponowany przez Friedman [4] jako metoda zespołowa umożliwiająca modelowanie złożonych, nielinowych zależności pomiędzy zmiennymi, co czyni go szczególnie przydatnym w zadaniach predykcyjnych opartych na danych regionalnych.

Dane źródłowe pozyskano z **Banku Danych Lokalnych Głównego Urzędu Statystycznego (BDL GUS)** i obejmowały informacje dotyczące liczby udzielonych noclegów oraz wybranych cech społeczno-gospodarczych powiatów w Polsce w latach 2013–2024. Na etapie przygotowania danych przeprowadzono proces czyszczenia zbiorów, ujednolicenia zakresów czasowych oraz inżynierii cech. W szczególności utworzono **zmienne opóźnione (lag features)**, umożliwiające prognozowanie wartości dla roku bieżącego na podstawie obserwacji z lat wcześniejszych.

W ramach badań porównano skuteczność czterech algorytmów regresyjnych powszechnie stosowanych w analizie predykcyjnej:

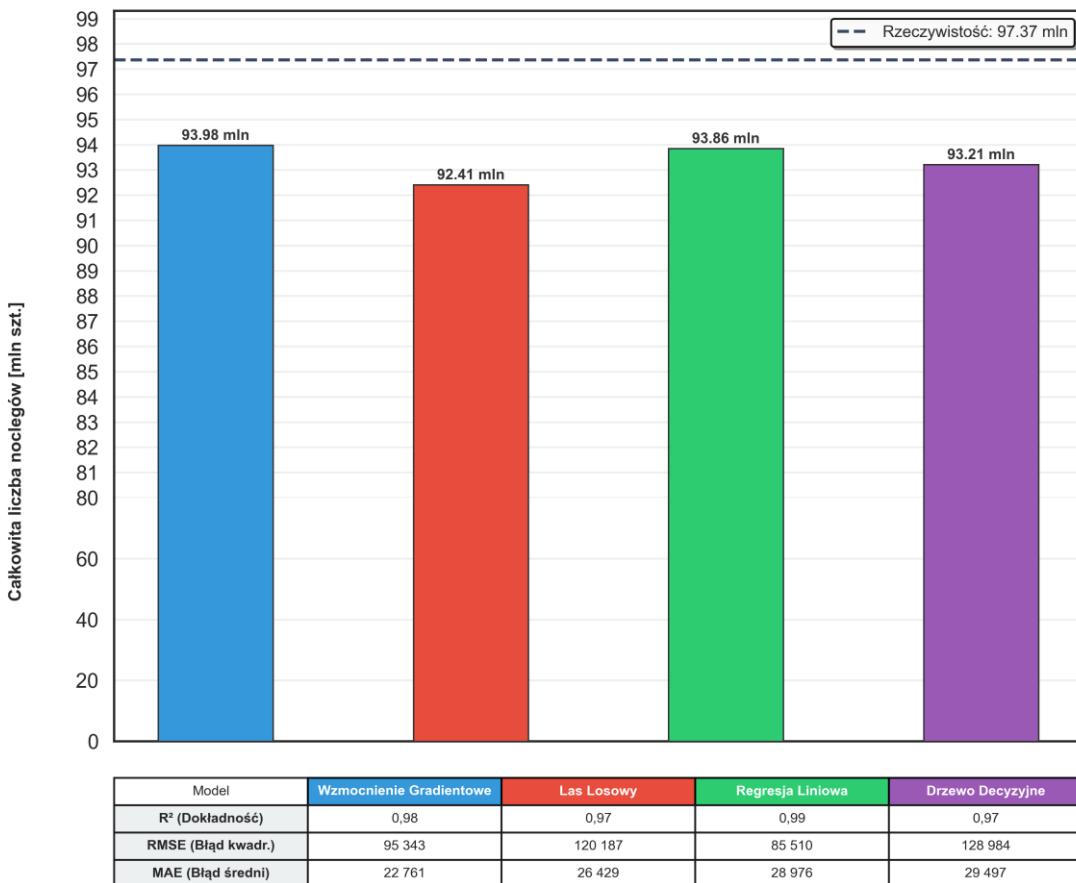
- **Regresji liniowej (Linear Regression)**,
- **Drzewa decyzyjnego (Decision Tree Regressor)**,
- **Lasu losowego (Random Forest Regressor)**,
- **Wzmocnienia gradientowego (Gradient Boosting Regressor)**.

4.1.1 Analiza porównawcza i ocena efektywności modeli regresyjnych

W celu wyłonienia optymalnego algorytmu predykcyjnego do prognozowania ruchu turystycznego przeprowadzono ewaluację porównawczą czterech modeli uczenia maszynowego: Wzmocnienie Gradientowe (Gradient Boosting), Las Losowy (Random Forest), Regresja Liniowa (Linear Regression) oraz Drzewo Decyzyjne (Decision Tree). Wybór tych metod podyktowany był ich powszechnym zastosowaniem w analizie szeregów czasowych i danych panelowych.

Jakość dopasowania modeli oceniono przy użyciu trzech kluczowych metryk statystycznych: współczynnika determinacji (R^2), błędu średniokwadratowego (RMSE) oraz średniego błędu bezwzględnego (MAE). Wyniki analizy przedstawiono na rycinie 1.

Ranking dokładności modeli (Ciągła skala zmienna)



Źródło: Opracowanie własne na podstawie danych BDL GUS

Ryc. 1. Ranking dokładności modeli regresyjnych w prognozowaniu liczby noclegów (2024) wraz z tabelą metryk błędów.

Źródło: *Opracowanie własne na podstawie danych BDL GUS.*

Analiza wizualna wykresu pozwala odnieść prognozy modeli do wartości referencyjnej, oznaczonej przerywaną linią (**rzeczywista liczba noclegów: 97,37 mln**). Jak widać, wszystkie badane algorytmy wykazały tendencję do nieznacznego niedoszacowania globalnego wolumenu, jednak różnice te mieszczą się w granicach błędu akceptowalnego dla skali makroekonomicznej.

Szczegółowe zestawienie wyników w tabeli wskazuje na zróżnicowaną efektywność algorytmów w zależności od przyjętego kryterium oceny:

- **Regresja Liniowa** osiągnęła najwyższy współczynnik determinacji **R² = 0,99** oraz najniższą wartość błędu średniokwadratowego **RMSE = 85510**. Jednakże wyższa wartość błędu bezwzględnego **MAE = 28976** w porównaniu do metody wzmocnienia gradientowego może sugerować tendencję do nadmiernego dopasowania (overfitting) w przypadku danych o charakterze nieliniowym.

- **Wzmocnienie Gradientowe (Gradient Boosting)**, mimo nieznacznie niższego współczynnika $R^2 = 0,98$ i wyższego błędu średniokwadratowego $RMSE = 95343$ w porównaniu do regresji liniowej, charakteryzowało się najniższym błędem średnim bezwzględnym wynoszącym $MAE = 22761$. Oznacza to, że model ten generował najmniejsze odchylenie od wartości rzeczywistych w przeliczeniu na pojedynczy powiat, będąc bardziej odpornym na wartości odstające w ujęciu lokalnym.
- **Modele oparte na drzewach decyzyjnych (Las Losowy i Drzewo Decyzyjne)** wykazały nieco niższą skuteczność $R^2 = 0,97$, generując wyższe wartości błędów RMSE (odpowiednio 120 187 i 128 984), co czyni je mniej precyzyjnymi w analizie zagregowanej.

Na podstawie analizy kryterium minimalizacji błędu **MAE**, jako finalny model prognostyczny wybrano **Gradient Boosting Regressor**. Algorytm ten zapewnia najlepszy balans pomiędzy dopasowaniem do trendu globalnego a precyją szacunków dla poszczególnych jednostek terytorialnych, co potwierdza jego użyteczność w badaniach nad turystyką.

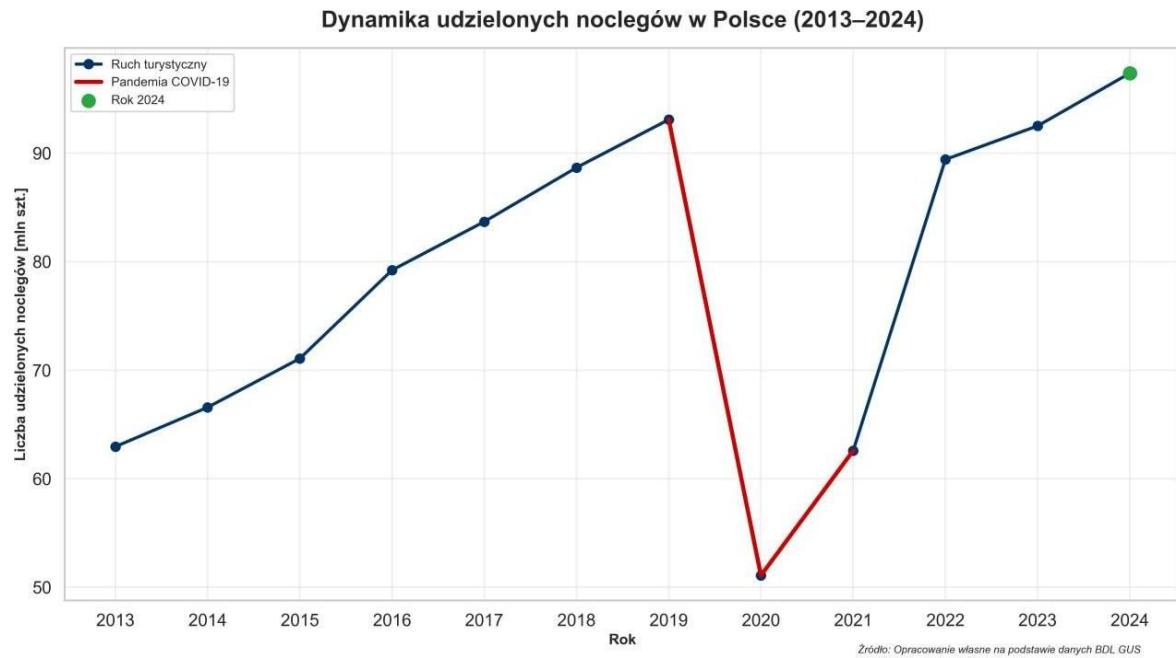
5. Analiza trendów ruchu turystycznego w latach 2013–2024

Na rycinie 2 przedstawiono zmiany liczby udzielonych noclegów w Polsce w analizowanym okresie. Analiza szeregu czasowego pozwala wyodrębnić kluczowe fazy cyklu koniunkturalnego w sektorze turystycznym.

W latach 2013–2019 obserwowano stabilny trend wzrostowy, skorelowany z ogólnym rozwojem gospodarczym kraju. W tym okresie liczba noclegów wzrosła z poziomu 62,9 mln do 93,3 mln, co wizualnie potwierdza rosnącą chłonność rynku turystycznego.

Gwałtowne załamanie trendu, widoczne w latach 2020–2021, nie było wynikiem czynników rynkowych, lecz bezpośrednim skutkiem wprowadzenia stanu epidemii na mocy stosownych regulacji prawnych [1] oraz restrykcji administracyjnych ograniczających działalność obiektów hotelarskich. W roku 2020 odnotowano gwałtowną redukcję liczby noclegów do poziomu 51,1 mln. Tak znaczący spadek, widoczny na wykresie jako ostre załamanie krzywej (oznaczone linią czerwoną), stanowi statystyczne potwierdzenie wrażliwości sektora na zewnętrzne szoki regulacyjne.

Okres po roku 2021 charakteryzuje się dynamiczną odbudową popytu. W 2024 roku prognozowana liczba noclegów osiągnęła wartość 97,4 mln, co oznacza powrót do tendencji wzrostowej i przekroczenie maksymalnych wartości odnotowanych przed wprowadzeniem ograniczeń.

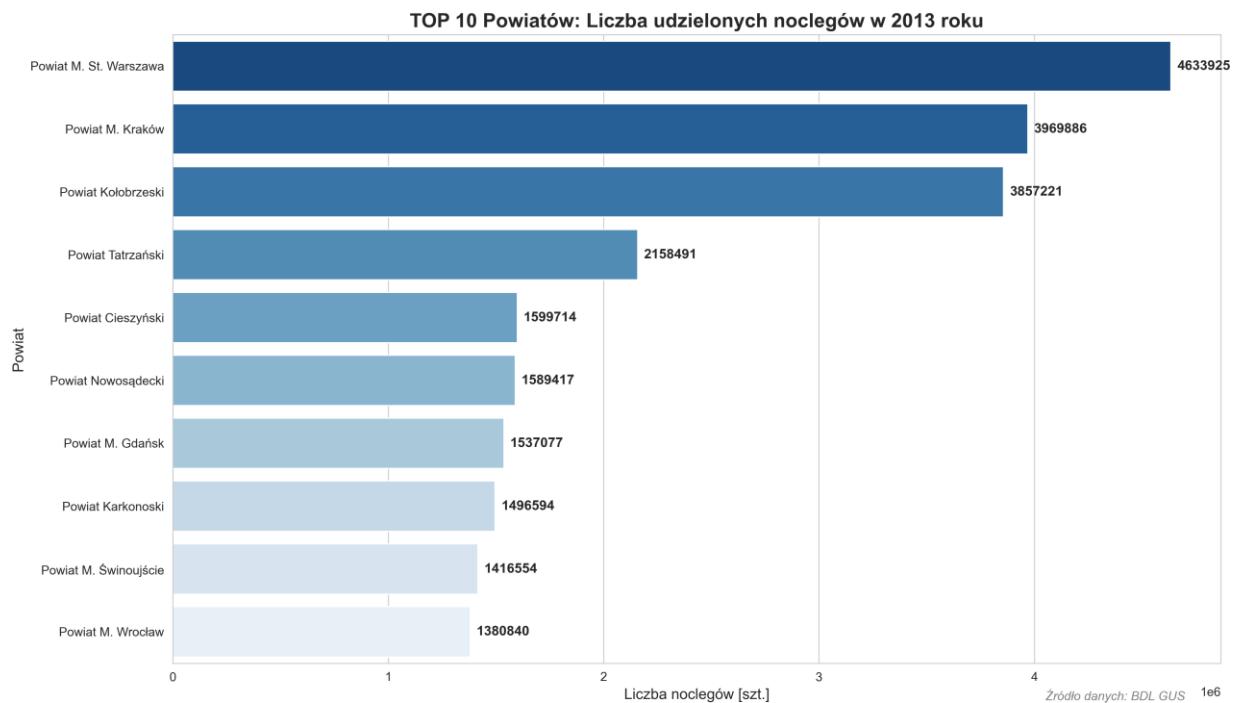


Ryc. 2. Dynamika zmian liczby udzielonych noclegów w Polsce w latach 2013–2024 z uwzględnieniem wpływu pandemii COVID-19. Źródło: *Opracowanie własne na podstawie danych BDL GUS*.

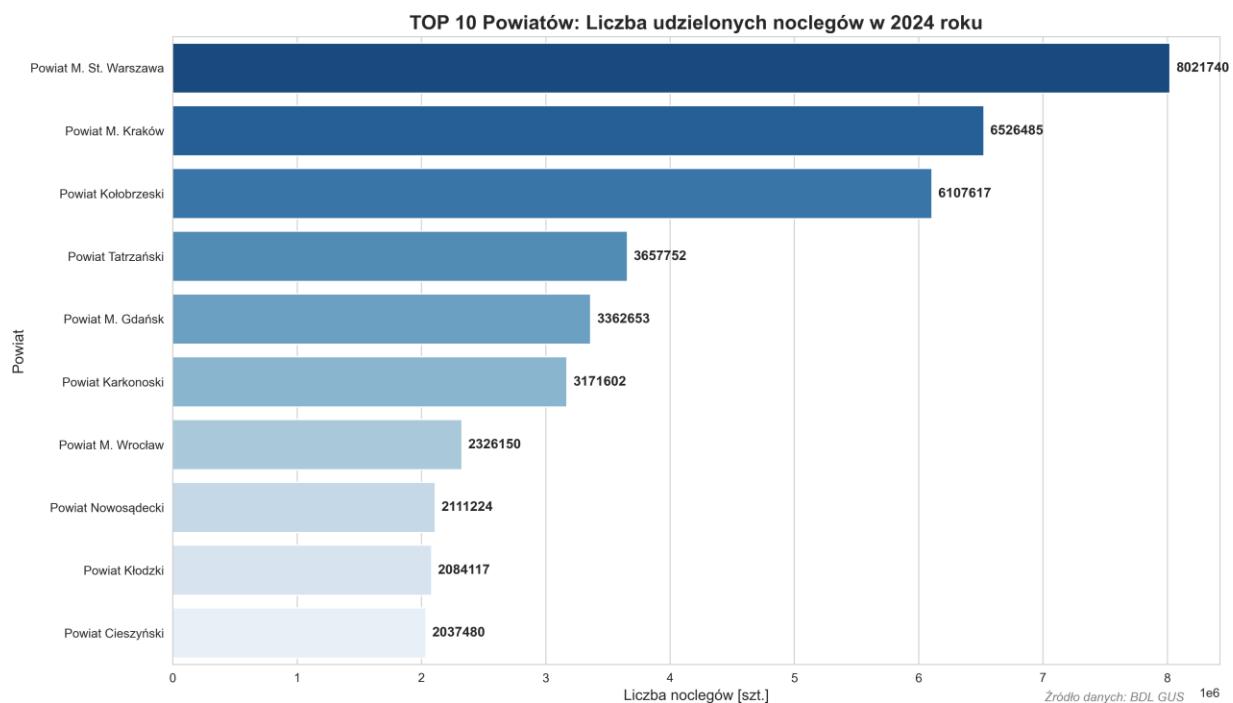
6. Zróżnicowanie przestrzenne i stabilność układu terytorialnego ruchu turystycznego

Analiza rozkładu przestrzennego liczby udzielonych noclegów w ujęciu powiatowym wskazuje na występowanie zjawiska silnej koncentracji geograficznej. Ruch turystyczny w Polsce nie rozkłada się równomiernie, lecz skupia w kluczowych jednostkach administracyjnych pełniących funkcję regionalnych hubów turystycznych. Zjawisko to jest zgodne z koncepcją przestrzeni turystycznej opisywaną w literaturze przedmiotu [2].

W celu oceny dynamiki zmian w strukturze przestrzennej przeprowadzono analizę komparatywną wiodących powiatów w skrajnych punktach badanego okresu, tj. w latach 2013 oraz 2024.



Ryc. 3.1. Struktura wiodących powiatów według liczby udzielonych noclegów w roku 2013.
Źródło: Opracowanie własne na podstawie danych BDL GUS.



Ryc. 3.2. Struktura wiodących powiatów według liczby udzielonych noclegów w roku 2024.
Źródło: Opracowanie własne na podstawie danych BDL GUS.

Analiza porównawcza powyższych zestawień (Ryc. 3.1 i Ryc. 3.2) pozwala sformułować następujące wnioski dotyczące ewolucji rynku noclegowego:

1. **Dominacja powiatów grodzkich:** Niezmiennym liderem krajowego rynku pozostaje Powiat m.st. Warszawa. W roku 2013 wolumen noclegów wynosił tam 4,63 mln,

natomiaszt w 2024 roku wzrósł do poziomu 8,02 mln, co oznacza przyrost o blisko 3,4 mln udzielonych noclegów. Podobną dynamikę wykazuje **Powiat m. Kraków**, gdzie odnotowano wzrost z 3,96 mln do 6,52 mln noclegów (przyrost o ponad 2,5 mln).

Wskazuje to na kluczową rolę powiatów będących dużymi ośrodkami miejskimi.

2. **Stabilność liderów wypoczynkowych:** W ścisłej czołówce stabilną pozycję utrzymują jednostki o profilu typowo turystycznym: **Powiat kołobrzeski** oraz **Powiat tatrzański**. Powiat kołobrzeski odnotował wzrost liczby noclegów z 3,85 mln (2013) do 6,10 mln (2024), umacniając swoją pozycję lidera w pasie nadmorskim. Z kolei w Powiecie tatrzańskim liczba noclegów wzrosła z 2,15 mln do 3,65 mln.
3. **Awans Powiatu m. Gdańsk:** Zauważalna jest istotna zmiana pozycji **Powiatu m. Gdańsk**, który w 2013 roku zajmował niższe miejsce z wynikiem 1,53 mln noclegów, a w 2024 roku awansował do ścisłej czołówki z wynikiem 3,36 mln, ponad dwukrotnie zwiększając swój potencjał noclegowy.

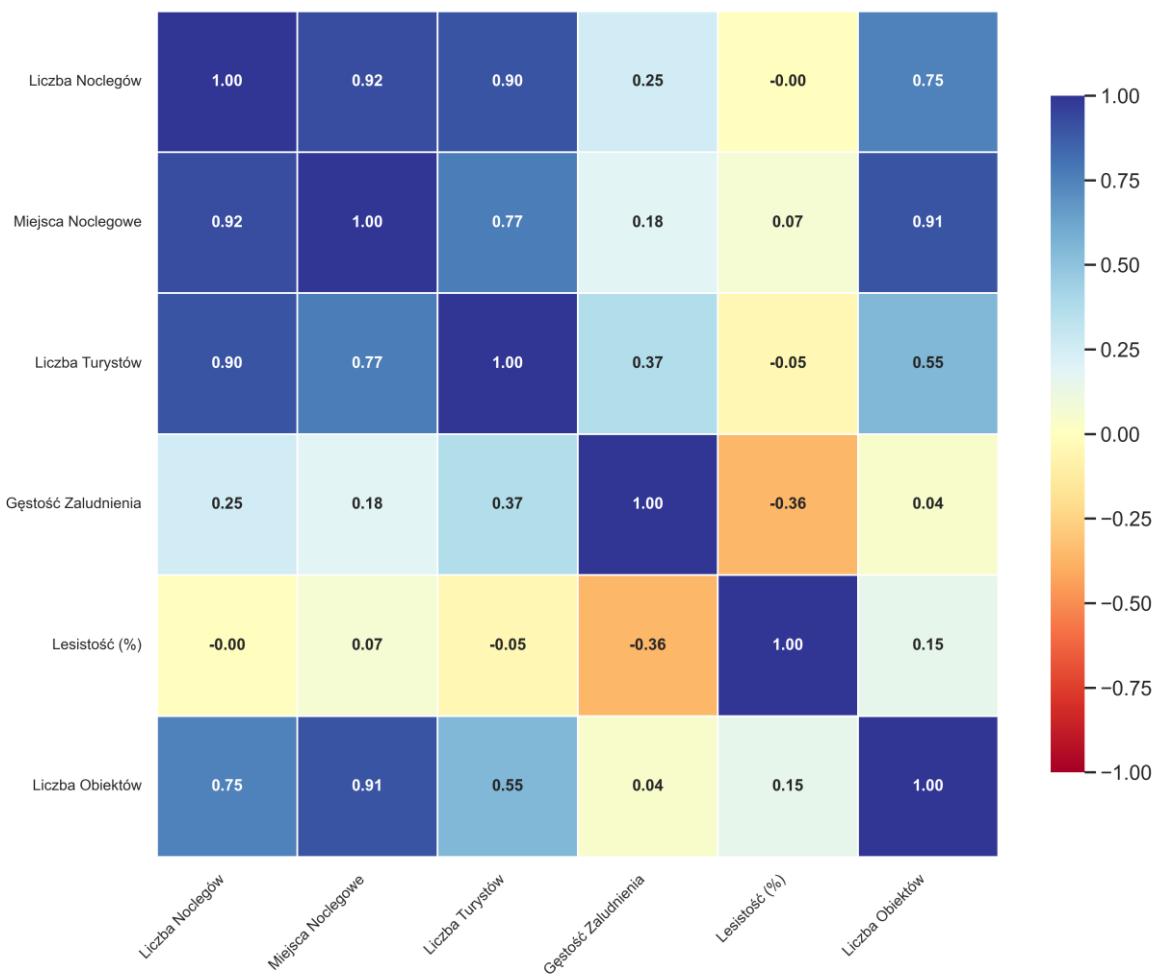
7. Analiza determinant popytu turystycznego – badanie korelacji

W celu identyfikacji kluczowych czynników wpływających na wielkość ruchu turystycznego (zmienna objaśniana: liczba udzielonych noclegów) przeprowadzono analizę korelacji liniowej Pearsona. Interpretację siły związku przeprowadzono w oparciu o skalę klasyfikacji współczynników korelacji według **M. Sobczyka [3]**, przyjmując, że $|r| \geq 0,9$ oznacza zależność bardzo silną, a $0,7 \leq |r| < 0,9$ – zależność silną.

Macierz korelacji (Ryc. 4) obrazuje siłę i kierunek zależności pomiędzy badanymi zmiennymi:

1. **Zależność bardzo silna:** Najsilniejszy związek z liczbą udzielonych noclegów wykazują **liczba miejsc noclegowych** ($r = 0,92$) oraz **liczba turystów** ($r = 0,90$). Zgodnie z przyjętą metodyką statystyczną [14], wartości te świadczą o niemal funkcyjnym powiązaniu podaży z popytem. Baza noclegowa jest zatem fundamentalną determinantą rozwoju – jej wielkość determinuje maksymalną chłonność turystyczną powiatu.
2. **Zależność silna:** Istotną statystycznie korelację odnotowano również w przypadku **liczby obiektów noclegowych** ($r = 0,75$). Niższa wartość w porównaniu do "miejsc noclegowych" wynika ze zróżnicowania wielkości obiektów (np. duży hotel vs mała agroturystyka).
3. **Brak istotnej zależności liniowej:** Zaskakującym wynikiem jest brak korelacji z **lesistością** ($r \approx 0,00$) oraz niska zależność z **gęstością zaludnienia** ($r = 0,25$). Wskazuje to, że w skali całego kraju walory przyrodnicze nie są wystarczającym stymulantem ruchu turystycznego bez odpowiedniej infrastruktury technicznej.

Macierz korelacji: Czynniki wpływające na turystykę



Źródło: Opracowanie własne na podstawie danych BDL GUS

Ryc. 4. Macierz korelacji Pearsona pomiędzy liczbą udzielonych noclegów a wybranymi wskaźnikami społeczno-gospodarczymi.

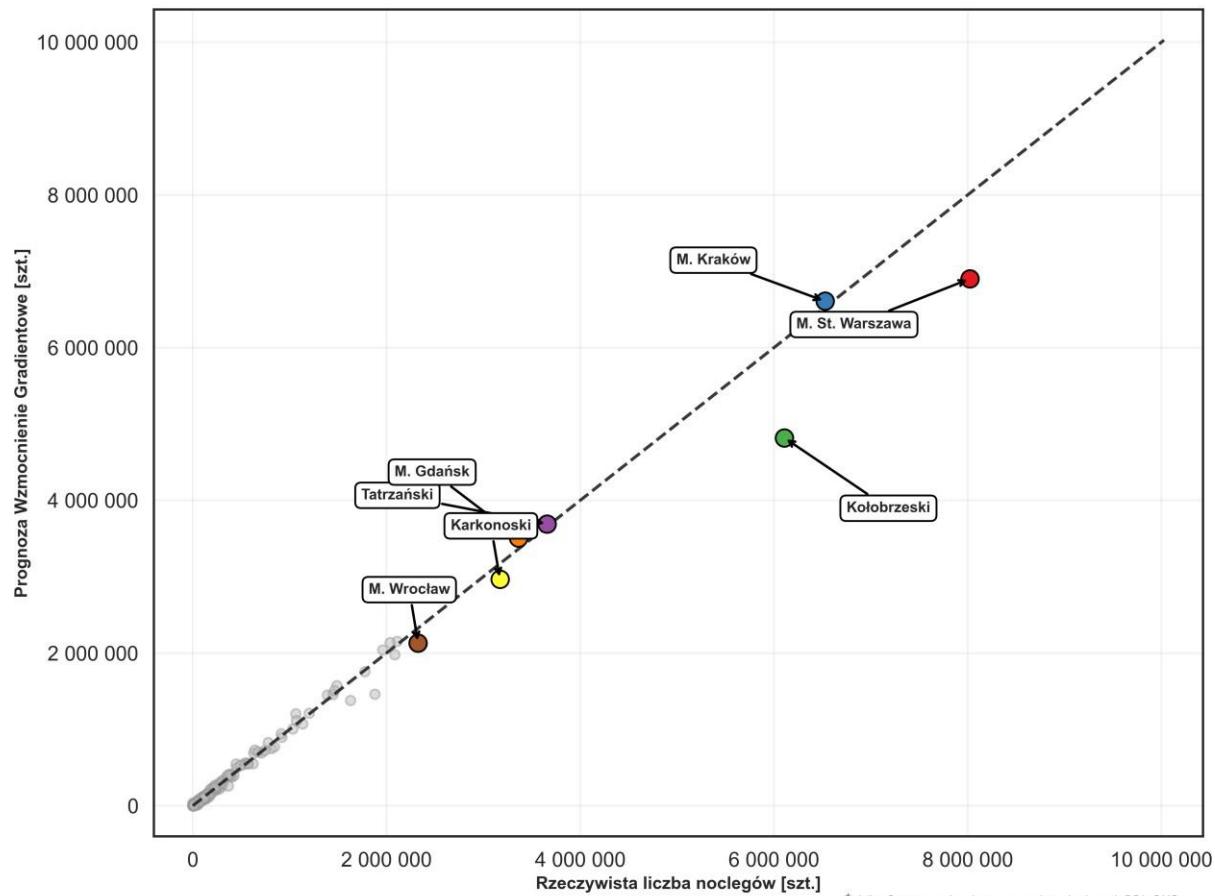
Źródło: Opracowanie własne na podstawie danych BDL GUS.

8. Walidacja modelu i ocena trafności prognoz lokalnych

Wybór finalnego modelu prognostycznego został poprzedzony szczegółową analizą porównawczą, przedstawioną w rozdziale 4. Na podstawie kryteriów statystycznych, takich jak współczynnik determinacji R^2 oraz błąd średni bezwzględny (MAE), jako narzędzie predykyjne wybrano algorytm **Wzmocnienia Gradientowego (Gradient Boosting)**. Metoda ta wykazała najwyższą zdolność generalizacji oraz najwyższy poziom błędów w skali ogólnokrajowej (szczegółowe zestawienie metryk błędów dla badanych modeli przedstawiono w tabeli na Ryc. 1).

W niniejszym rozdziale dokonano weryfikacji skuteczności wybranego modelu na danych rzeczywistych za rok 2024, koncentrując się na analizie odchyлеń w ujęciu lokalnym.

Wzmocnienie Gradientowe: Trafność prognozy (2024)



Źródło: Opracowanie własne na podstawie danych BDL GUS

Ryc. 5. Ocena trafności prognozy modelu Gradient Boosting: zestawienie wartości rzeczywistych i prognozowanych ($y=x$).

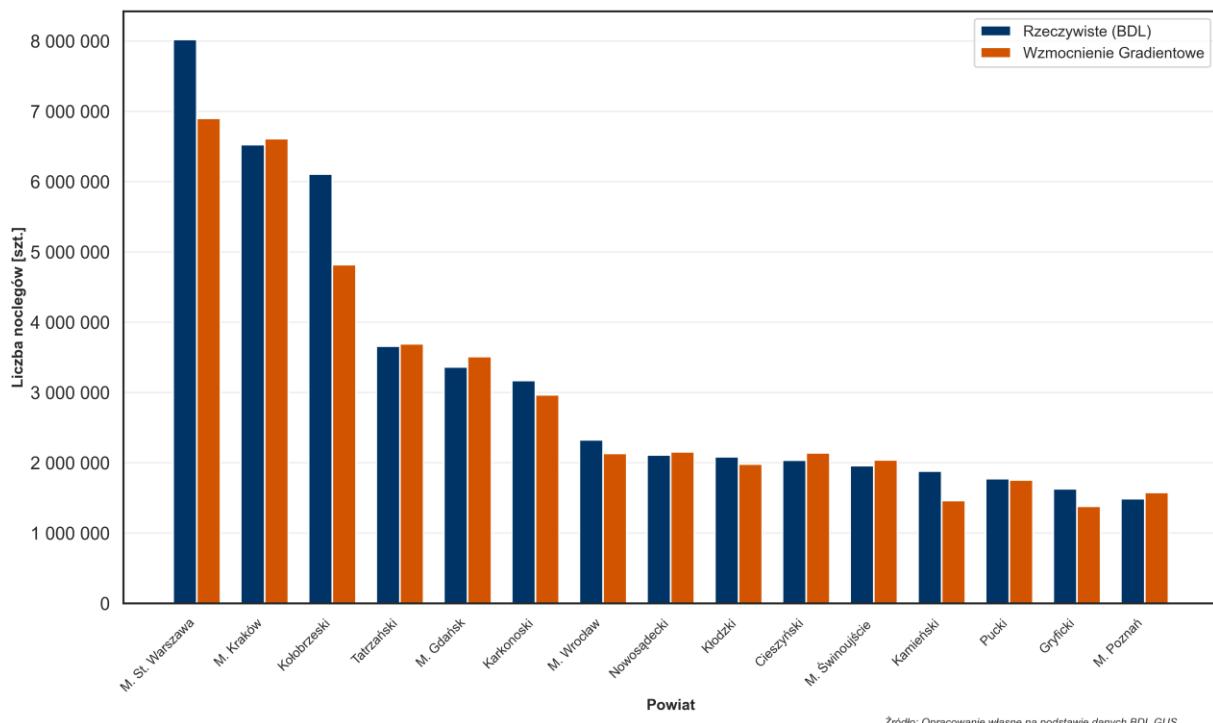
Źródło: *Opracowanie własne na podstawie danych BDL GUS.*

Analiza wykresu rozrzutu (Ryc. 5) pozwala na sformułowanie kluczowych wniosków dotyczących jakości dopasowania:

- Wysoka zbieżność globalna:** Punkty reprezentujące powiaty układają się wzduż przerywanej linii referencyjnej ($y=x$), co potwierdza wysoką precyzję modelu. Brak wyraźnych odchyleń systematycznych świadczy o tym, że model nie wykazuje tendencji do stałego zawyżania lub zaniżania prognoz dla ogółu zbioru.
- Identyfikacja liderów:** Model poprawnie wyodrębnił kluczowe ośrodki turystyczne (punkty w górnej prawej części wykresu: Warszawa, Kraków, Kołobrzeg), zachowując dystans względem pozostałych powiatów. Potwierdza to wysoki współczynnik determinacji ($R^2 \approx 0,98$) odnotowany w fazie testowej.

Szczegółową dekompozycję błędu dla 15 powiatów o największym ruchu turystycznym przedstawiono na wykresie słupkowym (Ryc. 6).

Wzmocnienie Gradientowe: Rzeczywistość vs Prognoza (2024)



Źródło: Opracowanie własne na podstawie danych BDL GUS

Ryc. 6. Porównanie wartości rzeczywistych i prognozowanych liczby noclegów w wiodących powiatach (2024).

Źródło: Opracowanie własne na podstawie danych BDL GUS.

Analiza porównawcza (Ryc. 6) ujawnia zróżnicowanie skuteczności modelu w zależności od specyfiki powiatu:

Specyfika metropolii (Warszawa): W przypadku lidera rankingu - m.st. Warszawy - widoczne jest wyraźne niedoszacowanie prognozy (słupek granatowy wyższy od pomarańczowego). Rzeczywisty ruch turystyczny (ok. 8 mln) przewyższył prognozę modelu (ok. 6,9 mln). Wynika to z faktu, że Warszawa generuje ogromny ruch biznesowy i dyplomatyczny, który wykracza poza standardowe ramy modelu opartego na bazie noclegowej.

Wysoka precyzja dla wiceliderów (Kraków, Tatry): W przypadku Krakowa oraz powiatu tatrzańskiego model wykazał się bardzo wysoką precyzją - słupki prognozowane i rzeczywiste osiągnęły niemal identyczne wartości. Oznacza to, że dla tych lokalizacji model Wzmocnienia Gradientowego działa optymalnie.

Niedoszacowanie w pasie nadmorskim (Kołobrzeg): W powiecie kołobrzeskim zauważalne jest niedoszacowanie wartości (słupek rzeczywisty wyższy od prognozy). Może to sugerować, że w kurortach nadmorskich występują dodatkowe czynniki przyciągające turystów (np. baza sanatoryjna, kwatery prywatne), które model wyceniał ostrożniej.

Poprawność hierarchii: Mimo różnic w wartościach bezwzględnych w Warszawie i Kołobrzegu, model poprawnie odwzorował kolejność (ranking) powiatów. Warszawa jest

pierwsza, Kraków drugi, a Kołobrzeg trzeci - zarówno w danych rzeczywistych, jak i w prognozie.

9. Prognozowanie przyszłego ruchu turystycznego

Opracowany w niniejszej pracy model predykyjny może zostać wykorzystany nie tylko do walidacji danych historycznych, lecz również do szacowania przyszłego poziomu ruchu turystycznego w polskich powiatach. Na podstawie relacji zaobserwowanych w danych z lat 2013–2024 możliwe jest generowanie prognoz liczby udzielonych noclegów dla kolejnych okresów czasowych, przy założeniu względnej stabilności analizowanych uwarunkowań społeczno-gospodarczych oraz infrastrukturalnych.

W szczególności zastosowany model Wzmocnienie Gradientowe (Gradient Boosting Regressor), który wykazał najwyższą trafność predykcyjną w walidacji lokalnej, stanowi solidną podstawę do dalszego prognozowania ruchu turystycznego.

Uzyskane prognozy wskazują na stopniową stabilizację oraz dalszy umiarkowany wzrost ruchu turystycznego w większości analizowanych jednostek, szczególnie w powiatach o rozwiniętej bazie noclegowej oraz ugruntowanej funkcji turystycznej. Jednocześnie należy podkreślić, że dokładność prognoz długookresowych jest wrażliwa na występowanie zdarzeń losowych, takich jak kryzysy gospodarcze, zmiany regulacyjne czy sytuacje nadzwyczajne o charakterze globalnym, analogiczne do pandemii COVID-19.

Z tego względu uzyskane wyniki prognostyczne należy traktować jako **scenariusz bazowy**, który może stanowić punkt odniesienia w procesach planowania rozwoju turystyki na poziomie regionalnym. Model może być w przyszłości rozwijany poprzez uwzględnienie dodatkowych zmiennych oraz aktualizację danych wejściowych, co pozwoli na dalsze zwiększenie jego użyteczności aplikacyjnej.

10. Ograniczenia badania i dalsze kierunki prac

Zastosowana metoda opiera się na danych rocznych, co uniemożliwia analizę sezonowości ruchu turystycznego. Ponadto model nie uwzględnia zdarzeń jednorazowych o charakterze lokalnym. W dalszych badaniach zasadne byłoby wykorzystanie danych o wyższej częstotliwości oraz zastosowanie alternatywnych metod modelowania.

Przykładem takiego podejścia mogłoby być wykorzystanie danych miesięcznych lub kwartalnych, co umożliwiłoby analizę sezonowości oraz krótkookresowych wahań popytu turystycznego.

11. Wnioski

Przeprowadzona analiza potwierdziła istotne zróżnicowanie przestrzenne ruchu turystycznego w polskich powiatach oraz kluczową rolę infrastruktury noclegowej w kształtowaniu liczby udzielonych noclegów. Najwyższe natężenie ruchu turystycznego odnotowano w dużych aglomeracjach miejskich oraz w regionach o ugruntowanej funkcji wypoczynkowej.

Wyniki modelowania wskazują, że dane Banku Danych Lokalnych GUS mogą być skutecznie wykorzystywane do prognozowania ruchu turystycznego na poziomie lokalnym. Spośród analizowanych algorytmów najwyższą trafność predykcyjną w ujęciu powiatowym wykazał model Gradient Boosting Regressor, który zapewnił najlepszy kompromis pomiędzy dokładnością prognoz a stabilnością wyników dla jednostek o wysokiej liczbie noclegów.

Uzyskane rezultaty potwierdzają zasadność wykorzystania zaawansowanych metod uczenia maszynowego w analizach regionalnych oraz wskazują na możliwość dalszego rozwoju modelu poprzez uwzględnienie dodatkowych zmiennych i aktualizację danych wejściowych wraz z pojawiением się nowych obserwacji statystycznych.

Bibliografia

- 1.[1] **Rozporządzenie Ministra Zdrowia z dnia 20 marca 2020 r. w sprawie ogłoszenia na obszarze Rzeczypospolitej Polskiej stanu epidemii (Dz.U. 2020 poz. 491). Dostęp: <http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20200000491>**
- 2.[2] Włodarczyk, B. (2014). *Przestrzeń turystyczna. Istota, koncepcje, determinanty rozwoju*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego
- 3.[3] Sobczyk, M. (2010). *Statystyka. Aspekty praktyczne i teoretyczne*. Warszawa: Wydawnictwo Naukowe PWN.
- 4.[4] Friedman, J. H. (2001). *Greedy Function Approximation: A Gradient Boosting Machine*. Annals of Statistics, 29(5), 1189–1232.
5. [5] Song, H., Li, G. (2008). Tourism demand modelling and forecasting – A review of recent research. *Tourism Management*, 29(2), 203–220.