

# Wykorzystanie **Big Data** w zastosowaniach transportowych na przykładzie danych pochodzących z automatów parkingowych<sup>2</sup>

**Artykuł przedstawia przykładową analizę podstawowych miar parkowania – poziomu napełnienia miejsc parkingowych w Strefie Płatnego Parkowania w Krakowie i rotacji korzystających z niej pojazdów, które uzyskano przy użyciu opracowanego przez autorkę algorytmu obliczeniowego zaimplementowanego w środowisku programistycznym Python. Przy użyciu metod regresji liniowej wielorakiej, stworzono modele predykcji poziomu napełnienia i rotacji w rejonach komunikacyjnych centralnej części miasta. Zmiennymi objaśniającymi modelu były podstawowe dane demograficzne pozyskane z modelu transportowego aglomeracji krakowskiej.**

## Wprowadzenie

W erze wszechobecnej cyfryzacji, każdego dnia, na niemal każdym kroku gromadzone są tzw. duże zestawy danych (ang. *Big Data*). Serwery zlokalizowane na całym świecie przechowują liczne ślady aktywności użytkowników telefonów komórkowych, internautów, klientów sklepów. Nieustannie dostarczają w ten sposób informacji pozwalających na obserwację aktualnych wyników sprzedażowych, czy położenia i przemieszczeń kierowców samochodów w czasie rzeczywistym. Posiadacze danych są natomiast w stanie lepiej dopasowywać się do dynamicznie zmieniającej się sytuacji na rynku czy drodze. Mogą przetwarzać liczne zgromadzone dane i na ich podstawie szukać przyczyn i analizować skutki przeszłych zdarzeń oraz tworzyć scenariusze i modele zdarzeń przyszłych.

Wiadomości zawarte w dużych bazach danych mogą być wykorzystywane zarówno przez liczne przedsiębiorstwa prywatne, jak i jednostki administracyjne państw oraz samorządów. Już dziś mówi się, że używanie *Big Data* stanowi przełom w zarządzaniu firmą, pozwalając na tworzenie nowych modeli biznesowych [6]. Powstawanie zdigitalizowanych, ogromnych zbiorów informacji, stwarza niepotykane dotychczas warunki dla osób badających zjawiska z wielu dziedzin. Obok sektorów takich, jak marketing czy bankowość, znajduje się również szeroko pojęta inżynieria logistyki i transportu.

## Przegląd literatury

Pozyskanie danych oraz rola ich właściwej interpretacji jest jednym z najistotniejszych wyzwań współczesnej nauki, także w dziedzinie transportu [4, 9]. Przez setki lat dane były gromadzone i archiwizowane w formie papierowej, stanowiąc dokumentację zdarzeń czy transakcji, które miały miejsce w przeszłości. Obecnie, przede wszystkim dzięki cyfryzacji danych, wszechobecnemu dostępowi do sieci Internet oraz wystarczającej pojemności serwerów, możliwy stał się rozkwit technologii związanych z analizą *Big Data*. Źródłem *Big Data* w zastosowaniach transportowych jest bardzo wiele. Mogą być nimi zarówno systemy rowerów miejskich, rejestry dostarczonych przesyłek kurierskich, jak i aplikacje korporacji taksówkarskich.

Przykładowym zastosowaniem *Big Data*, na jakie można natrafić w literaturze, jest m.in. zarządzanie łańcuchem dostaw [7]. Przedsiębiorstwa są w stanie – na podstawie analizy dużych danych – na bieżąco optymalizować trasy dojazdu, szczegółowo planować zdolności operacyjne, dokonywać kompleksowych analiz ryzyka itp. Kolejnym ze sposobów wykorzystania *Big Data* do celów szeroko pojętych analiz transportowych jest również tworzenie macierzy przemieszczeń na podstawie danych pochodzących od operatorów komórkowych [1]. Macierze te stanowią jedno z możliwych źródeł danych wejściowych do modeli transportowych. Modele te mogą być zaś

<sup>1</sup> Mgr inż. C. Kubala, doktorantka, Politechnika Krakowska, Wydział Inżynierii Lądowej, Instytut Inżynierii Drogowej, Kolejowej i Transportu, Katedra Systemów Transportowych.

<sup>2</sup> Artykuł recenzowany.

kalibrowane, na kolejnych etapach prac, do pomiarów prędkości uzyskanych z masowo gromadzonych danych GPS [5]. Innym przykładem transportowego zastosowania *Big Data* są próby predykcji (w czasie rzeczywistym) poziomów napełnień publicznie dostępnych miejsc parkingowych. Do analizy użyto danych jednego z operatorów telefonii komórkowej, za pośrednictwem którego możliwe było uiszczenie opłaty za postój [3].

Wspomniane wyżej przykłady, zaczerpnięte z literatury, pozwoliły na identyfikację zróżnicowanych zastosowań *Big Data* w zagadnieniach transportowych. Stanowią one jednak zaledwie niewielki odsetek możliwości oferowanych przez tzw. duże dane. Zarówno zakres, jak i tematyka analiz jest, przy posiadaniu dużych zbiorów danych, limitowana w największej mierze przez ograniczenia ludzkiej percepcji i wyobraźni.

### Model zachowań parkingowych kierowców

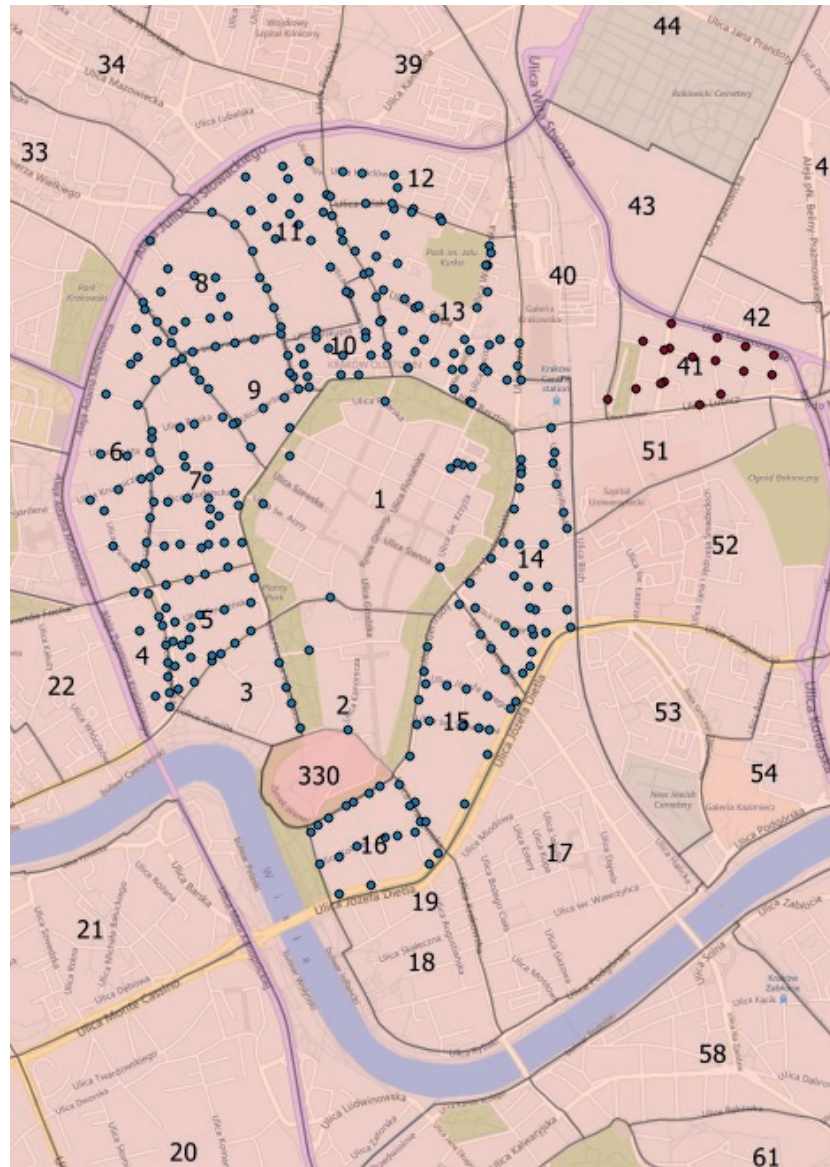
W artykule podjęto próbę analizy *Big Data* gromadzonych w krakowskiej Strefie Płatnego Parkowania przy pomocy miejskich parkomatów. Zdecydowano się na użycie przykładowych danych otrzymanych dzięki uprzejmości zarządcy tej strefy, komunalnej firmy Miejska Infrastruktura Sp. z o.o. Pozyskano informacje zbierane od momentu uruchomienia pierwszych automatów parkingowych (lipiec 2009 roku) do dnia ekstrakcji danych (24 maja 2016 roku). Baza danych zawierała informacje dla obszarów, które w dniu instalacji parkomatów znajdowały się w Strefie Płatnego Parkowania ścisłego centrum Krakowa. Otrzymano także informacje z Podstrefy P3 (znajdującej się w bezpośrednim sąsiedztwie dworca kolejowego i autobusowego), która została włączona do Strefy Płatnego Parkowania w późniejszym czasie. Lokalizacje parkomatów uwzględnionych w analizie przedstawiono na rysunku 1.

Każdy z rekordów (tj. wierszy) bazy danych zawierał informacje odnośnie pojedynczego wydanej biletu parkingowego, czyli:

- numer identyfikacyjny parkomatu, z którego został pobrany
- godzinę i datę pobrania
- czas, na jaki bilet parkingowy został wykupiony.

Jednakże, ze względu na obszerny horyzont czasowy dostępnej bazy danych parkomatowych (zajmującej ponad 16,5 mln rekordów) postanowiono opracować dane jedynie dla jednego wybranego miesiąca funkcjonowania Strefy Płatnego Parkowania (wybrano maj 2015 roku). Uszczuplenie zakresu danych spowodowało utworzenie podzbioru bazy danych liczącego 353 469 rekordów.

Dysponując tak szczegółowymi informacjami, postanowiono spróbować uchwycić zależność między



Rys. 1. Lokalizacje automatów parkingowych, z których udostępnione zostały dane dotyczące wykupionych biletów parkingowych.

Źródło: opracowanie własne na podstawie [8].

zachowaniami parkingowymi kierowców, a podstawowymi charakterystykami społeczno-gospodarczymi poszczególnych obszarów obowiązywania Strefy Płatnego Parkowania. Do stworzenia modeli odwzorowujących zachowania parkingowe użytkowników miejskiej infrastruktury postanowiono wykorzystać dwie najbardziej istotne miary wykorzystania parkingów i warunków parkowania, tj. poziom napełnień oraz stopień rotacji pojazdów [2]. Pierwszy z uwzględnianych wskaźników rozumiany jest jako procentowy udział liczby zaparkowanych pojazdów w całkowitej liczbie dostępnych miejsc parkingowych na danym obszarze w jednostce czasu. Wskaźnik rotacji wyraża liczbę pojazdów, jakie skorzystały z jednego miejsca postojowego na danym obszarze w jednostce czasu. Przyjętą jednostką czasu dla poziomu napełnień była godzina, rotację zaś postanowiono liczyć w odniesieniu do doby. Dane agregowane były do poziomu rejonów komu-

nikacyjnych pozyskanych z modelu transportowego miasta [8]. Rejony te charakteryzują się podobnymi zachowaniami transportowymi, przez co możliwe jest osadzenie uwarunkowań społeczno-ekonomicznych w kontekście transportowym.

Obliczenia obu miar parkowania dla każdego z rejonów komunikacyjnych dokonano na podstawie analizy bazy danych wykonanej przy użyciu autorskiego algorytmu zaimplementowanego w środowisku programistycznym Python oraz dzięki znajomości dokładnej lokalizacji każdego z parkomatów. Modele matematyczne zostały natomiast stworzone przy użyciu metod regresji wielorakiej liniowej. Zmiennymi objaśnianymi były poziomy napętnień w każdej z dziesięciu godzin obowiązywania Strefy Płatnego Parkowania oraz dobowy stopień rotacji (łącznie 11 modeli). Przyjętymi zmiennymi objaśniającymi były zawsze wyłącznie dane wejściowe do modelu transportowego, z którego zaczerpnięty został podział na rejony komunikacyjne miasta. Jako zmienne objaśniające uwzględniono liczbę dostępnych miejsc parkingowych, powierzchnię produkcyjną, powierzchnię przemysłową oraz liczbę mieszkańców. W charakterze zmiennych objaśniających nie wykorzystano wyników rozkładu modelu ruchu na sieć. Poziom ufności przyjęto na poziomie  $\alpha = 0.05$ , co jest poziomem typowym dla tego rodzaju badań. Równania użyte do budowy modeli przedstawiono poniżej:

$$\text{Napętnienie}_{i,j} = a_{0,j} + a_{1,j} * X_{1,i} + a_{2,j} * X_{2,i} + a_{3,j} * X_{3,i} + a_{4,j} * X_{4,i} \quad (1)$$

$$\text{Rotacja}_i = b_0 + b_1 * X_{1,i} + b_2 * X_{2,i} + b_3 * X_{3,i} + b_4 * X_{4,i} \quad (2)$$

gdzie:

$\text{Napętnienie}_{i,j}$  – prognozowany poziom napętnienia i-tego rejonu komunikacyjnego w j-tej godzinie obowiązywania Strefy Płatnego Parkowania [%],

$\text{Rotacja}_i$  – prognozowany dobowy poziom rotacji w i-tym rejonie komunikacyjnym,

$$a_{0,j}; a_{1,j}; a_{2,j}; a_{3,j}; a_{4,j}$$

– parametry modelu regresji napętnień dla j-tej godziny,

$$b_0; b_1; b_2; b_3; b_4$$

– parametry modelu regresji dobowej rotacji,

$X_{1,i}$  – liczba dostępnych miejsc parkingowych w i-tym rejonie komunikacyjnym,

$X_{2,i}$  – powierzchnia produkcyjna w i-tym rejonie komunikacyjnym [m<sup>2</sup>],

$X_{3,i}$  – powierzchnia przemysłowa w i-tym rejonie komunikacyjnym [m<sup>2</sup>],

$X_{4,i}$  – liczba mieszkańców i-tego rejonu komunikacyjnego.

### Wyniki oraz analiza badań

Dzięki powiązaniu informacji pochodzących z bazy danych parkingowych, z charakterystyką społeczno-gospodarczą poszczególnych rejonów komunikacyjnych, możliwe było obliczenie parametrów modeli wyznaczających godzinowy poziom napętnień i dobowej rotacji w krakowskiej Strefie Płatnego Parkowania. Kalkulacji modeli dokonano za pomocą metody najmniejszych kwadratów, przy użyciu dedykowanego oprogramowania statystycznego. Szczegółowe wyniki analizy przedstawione zostały w tabelach 1 i 2.

W przypadku analizy regresji przeprowadzonej dla poziomu napętnień w poszczególnych godzinach obowiązywania Strefy Płatnego Parkowania udało się uzyskać wartości współczynnika determinacji  $R^2$  w przedziale od 0.63 do 0.83 (średnio 0.74). Ponadto, sąsiednie wartości wskaźników ułożonych chronologicznie, są do siebie bardzo zbliżone i brak jest skokowych zmian. Ze statystyk wynika, że użyte zmienne objaśniające najlepiej tłumaczą poziom napętnień poszczególnych rejonów komunikacyjnych w godzinach wczesnopopołudniowych, określanych często

Tab. 1. Parametry dla modelu wyznaczającego prognozowany godzinowy poziom napętnień w Strefie Płatnego Parkowania przez pojazdy wykupujące bilety parkingowe.

	Godzina									
	10	11	12	13	14	15	16	17	18	19
$a_0$	44.0	41.3	38.0	36.6	34.8	32.0	30.5	32.4	34.6	33.7
$a_1$	-0.070	-0.068	-0.064	-0.062	-0.059	-0.054	-0.049	-0.052	-0.056	-0.056
$a_2$	0.0054	0.0053	0.0051	0.0050	0.0045	0.0039	0.0031	0.0030	0.0038	0.0046
$a_3$	-0.0039	-0.0037	-0.0036	-0.0035	-0.0032	-0.0028	-0.0024	-0.0025	-0.0030	-0.0033
$a_4$	0.0088	0.0087	0.0084	0.0081	0.0077	0.0072	0.0065	0.0069	0.0074	0.0078
$R^2$	<b>0.77</b>	<b>0.79</b>	<b>0.82</b>	<b>0.83</b>	<b>0.82</b>	<b>0.77</b>	<b>0.70</b>	<b>0.66</b>	<b>0.65</b>	<b>0.63</b>

Źródło: opracowanie własne

mianem tzw. „międzyszczytu” komunikacyjnego. Najbardziej estymowane były natomiast napętnienia dla godzin późnego popołudnia, w związku z czym muszą istnieć nieuwzględnione w modelu czynniki, które wpływają na obserwowane zjawisko zwłaszcza o tej porze.

**Tab. 2. Parametry dla modelu wyznaczającego prognozowany dobowy poziom rotacji w Strefie Płatnego Parkowania dla pojazdów wykupujących bilety parkingowe.**

Współczynnik	
$b_0$	2.734347
$b_1$	-0.004258
$b_2$	0.000277
$b_3$	-0.000222
$b_4$	0.000532

Źródło: opracowanie własne

W przypadku próby zamodelowania dobowego stopnia rotacji, współczynnik determinacji  $R^2$  wyniósł 0.77. Porównując uzyskaną wartość  $R^2$  modelu dla rotacji z wynikami uzyskanymi dla prognoz poziomu napętnień, wskaźniki wydają się być bardzo zbliżone. Pomimo tego, że napętnienia godzinowe nie implikują stopnia rotacji, to wszystkie odpowiadające sobie parametry (np.  $a_{0,10} - b_0$ ,  $a_{3,13} - b_3$ , ale też  $a_{3,13} - a_{3,17}$ ) zachowywały te same znaki. Wobec powyższego, wyniki mogą świadczyć o bardzo dobrej jakości (spójności) danych, co przekłada się bezpośrednio na wysoki potencjał wykorzystania naukowego parkingowych *Big Data*, jakimi są informacje pochodzące z automatów parkingowych.

## Wnioski

Analiza *Big Data* stanowi bez wątpienia przyszłość inżynierii transportu. Dzieje się tak m.in. za sprawą łatwości ich pozyskania, możliwości gromadzenia informacji dla szerokich horyzontów czasowych oraz, przede wszystkim, dzięki istnieniu zoptymalizowanych narzędzi informatycznych, które pozwalają te dane obrabiać. W artykule wykazano, że nawet przy niewielkiej liczbie użytych zmiennych objaśniających, istnieje możliwość uzyskania wskaźników determinacji poszczególnych modeli regresji na zadowalającym poziomie. Wyniki otrzymane już w ramach pierwszych badań, dla zawężonego okresu analizy, zachęcają do kontynuacji poszukiwań prowadzących do uzupełnienia zestawu zmiennych wyjaśniających obserwowane zjawisko.

Przykład przedstawionego w artykule opracowania może (i powinien) stanowić zachętę do podejmowania prób poszukiwania odpowiedzi na nurtujące pytania właśnie w analizie dużych baz danych.

Warto zaznaczyć, iż zachęta ta nie dotyczy jedynie osób prowadzących badania naukowe, lecz także specjalistów pionów technicznych i zarządczych różnych branż.

### Streszczenie

W artykule postanowiono podjąć próbę wykorzystania tzw. *Big Data* do celów analizy zachowań parkingowych użytkowników krakowskiej Strefy Płatnego Parkowania. W ramach badań pokazano, jak odpowiednia agregacja i przetworzenie informacji pochodzących z bazy danych automatów parkingowych pozwala na identyfikację zachowań kierowców. Artykuł przedstawia przykładową analizę podstawowych miar parkowania – poziomu napętnień i rotacji, które zostały uzyskane przy użyciu autorskiego algorytmu obliczeniowego zaimplementowanego w środowisku programistycznym Python. Przy użyciu metod regresji liniowej wielorakiej, stworzono modele predykcji poziomu napętnień i rotacji w rejonach komunikacyjnych centralnej części miasta. Zmiennymi objaśniającymi modelu były podstawowe dane demograficzne pozyskane z modelu transportowego aglomeracji krakowskiej.

**Słowa kluczowe:** Big Data, transport, parkowanie, rotacja pojazdów.

## Use of Big Data in transport applications on the basis of data from parking machines

### Abstract

In the article it was decided to make an attempt to use the so-called Big Data for the purpose of analysis of parking behaviour of users of the Krakow Paid Parking Zone (Poland). The research showed how appropriate aggregation and processing of information from the parking machine database allows for identification of drivers' behaviour. The article presents a sample analysis of basic parking measures - the level of occupancy and rotation, which were obtained using an original calculation algorithm implemented in the Python development environment. Using multiple linear regression methods, prediction models of the level of occupancy and rotation in traffic analysis zones of the central part of the city were developed. The variables explaining the model were the basic demographic data obtained from the transport model of the Kraków agglomeration.

**Keywords:** Big Data, transport, parking, rotation of vehicles.

### LITERATURA/BIBLIOGRAPHY

- [1] Caceres, N. i in. (2012) Traffic Flow Estimation Models Using Cellular Phone Data, *Intelligent Transportation Systems, IEEE Transactions* 13, ss. 1430-1441.
- [2] Gaca, S., Suchorzewski, W. i Tracz, M. (2011) *Inżynieria ruchu drogowego: teoria i praktyka*. Warszawa: Wydawnictwo Komunikacji i Łączności.
- [3] Hoessinger, R. i in. (2014) „Development of a Real-Time Model of the Occupancy of Short-Term Parking Zones”, *International Journal of Intelligent Transportation Systems Research*, ss. 37-47.
- [4] Jin X. i in. (2015) Significance and Challenges of Big Data Research, *Big Data Research, Vol 2 Issue 2*, ss. 59 – 64.
- [5] Herrera J. C. i in. (2009) Evaluation of Traffic Data Obtained via GPS-enabled Mobile Phones: The Mobile Century field experiment, *UCB-ITS-VWP-2009-8, Berkeley Center for Future Urban Transport*, Uniwersytet Kalifornijski.
- [6] McAfee, A., Brynjolfsson, E. (2013) Big Data, czyli przełom w zarządzaniu firmą, *Harvard Business Review Polska*, 4, ss. 42 – 51.
- [7] Marciniak M., Szymczak M. (2015) Big data w zarządzaniu łańcuchem dostaw, *Gospodarka Materiałowa i Logistyka*, nr 7/2015.
- [8] Pracownia Badań Społecznych i in. (2014) *Badania zachowań komunikacyjnych mieszkańców Krakowskiego Obszaru Metropolitalnego*, Zleceniodawca: Urząd Miasta Krakowa, Wydział Gospodarki Komunalnej i Ochrony Środowiska.
- [9] Szarata, A. (2015) Rola pozyskiwania danych w kontekście funkcjonowania Stref Płatnego Parkowania, *Zeszyty Naukowo – Techniczne Oddziału SITK w Krakowie, 1(105)/2015*, ss. 107-117.