

Politechnika Warszawska

WYDZIAŁ ELEKTRONIKI  
I TECHNIK INFORMACYJNYCH



Instytut Automatyki i Informatyki Stosowanej

# Praca dyplomowa magisterska

na kierunku Informatyka  
w specjalności Systemy Informacyjno-Decyzyjne

Modelowanie ruchu drogowego na podstawie danych rzeczywistych

**Łukasz Fijas**

Numer albumu 229440

promotor  
dr inż. Mariusz Kamola

Warszawa 2019



## Streszczenie

Celem niniejszej pracy magisterskiej było opracowanie modeli umożliwiających symulację ruchu drogowego w sposób jak najbliższy rzeczywistym zachowaniom kierowców. W dążeniu do znalezienia cech charakterystycznych i wzorców niezbędnych do modelowania ruchu pojazdów, szczegółowej analizie poddane zostały dostępne dane historyczne. Zbadane zostały zagadnienia zgodności rzeczywistych tras przejazdów z wyznaczonymi trasami optymalnymi, wpływu informacji o aktualnej sytuacji na drogach na decyzje kierowców oraz skutków zmian topologii sieci dróg na rozkład natężenia ruchu. W pracy porównano również działanie wybranych modeli klasyfikacyjnych na przykładzie kategoryzacji typów kierowców.

**Słowa kluczowe:** inżynieria ruchu drogowego, modelowanie ruchu drogowego, metody klasyfikacji, dane drogowe czasu rzeczywistego, OpenStreetMap, GraphHopper

## Abstract

### Traffic modelling based on real data

The aim of the master's thesis was to develop models, which can be used for simulating traffic as similar as possible to the routes chosen by real drivers. Available historical traffic data was deeply analyzed in order to determine patterns needed while elaborating the models. The following aspects have been studied: coverage of real routes with calculated optimal routes, influence of real time traffic information on drivers' decisions and effects of road network modifications on traffic flow. The thesis also contains the comparison of selected classification models illustrated by the example of drivers classification.

**Keywords:** traffic engineering, traffic modelling, classification methods, real time traffic data, OpenStreetMap, GraphHopper





„załącznik nr 3 do zarządzenia nr 24/2016 Rektora PW

.....  
miejscowość i data

.....  
imię i nazwisko studenta

.....  
numer albumu

.....  
kierunek studiów

### OŚWIADCZENIE

Świadomy/-a odpowiedzialności karnej za składanie fałszywych zeznań oświadczam, że niniejsza praca dyplomowa została napisana przeze mnie samodzielnie, pod opieką kierującego pracą dyplomową.

Jednocześnie oświadczam, że:

- niniejsza praca dyplomowa nie narusza praw autorskich w rozumieniu ustawy z dnia 4 lutego 1994 roku o prawie autorskim i prawach pokrewnych (Dz.U. z 2006 r. Nr 90, poz. 631 z późn. zm.) oraz dóbr osobistych chronionych prawem cywilnym,
- niniejsza praca dyplomowa nie zawiera danych i informacji, które uzyskałem/-am w sposób niedozwolony,
- niniejsza praca dyplomowa nie była wcześniej podstawą żadnej innej urzędowej procedury związanej z nadawaniem dyplomów lub tytułów zawodowych,
- wszystkie informacje umieszczone w niniejszej pracy, uzyskane ze źródeł pisanych i elektronicznych, zostały udokumentowane w wykazie literatury odpowiednimi odnośnikami,
- znam regulacje prawne Politechniki Warszawskiej w sprawie zarządzania prawami autorskimi i prawami pokrewnymi, prawami własności przemysłowej oraz zasadami komercjalizacji.

Oświadczam, że treść pracy dyplomowej w wersji drukowanej, treść pracy dyplomowej zawartej na nośniku elektronicznym (płycie kompaktowej) oraz treść pracy dyplomowej w module APD systemu USOS są identyczne.

.....  
czytelny podpis studenta”



# Spis treści

<b>1</b>	<b>Wstęp</b>	<b>9</b>
1.1	Wprowadzenie i cel pracy . . . . .	9
1.2	Przegląd aktualnego stanu badań i dostępnych rozwiązań . . .	10
1.2.1	Paradoks Braessa . . . . .	10
1.2.2	Informacja o aktualnym natężeniu ruchu drogowego kiedyś a dziś . . . . .	11
1.2.3	Analiza i inżynieria ruchu na podstawie ankiet . . . . .	12
1.2.4	Modelowanie ruchu drogowego . . . . .	13
1.2.5	Informacje w czasie rzeczywistym a płynność ruchu . . .	14
1.2.6	Modele predykcji ruchu a zachowanie kierowców . . . . .	15
1.2.7	Samochodowe systemy GPS a modelowanie ruchu . . . . .	16
1.2.8	Efektywne modelowanie ruchu drogowego w czasie rze- czywistym . . . . .	16
<b>2</b>	<b>Stworzenie środowiska analitycznego</b>	<b>19</b>
2.1	Przegląd wykorzystywanych narzędzi . . . . .	19
2.2	Środowisko analityczne . . . . .	21
2.2.1	Definicje wykorzystywanych pojęć . . . . .	21
2.2.2	Wyznaczanie tras optymalnych . . . . .	22
2.2.3	Wyznaczanie części wspólnych tras . . . . .	24
2.2.4	Modelowanie natężenia ruchu . . . . .	27
2.2.5	Wykrywanie postojów . . . . .	28
<b>3</b>	<b>Analiza danych</b>	<b>31</b>
3.1	Czy kierowcy jeżdżą trasami optymalnymi? . . . . .	31
3.1.1	Analiza przejazdów bez postojów . . . . .	31
3.1.2	Długość trasy a pokrycie z trasą optymalną . . . . .	32
3.1.3	Trasa optymalna – najkrótsza czy najszybsza? . . . . .	32
3.2	Analiza uwzględniająca natężenie ruchu . . . . .	33
3.3	Dlaczego miałbym zrezygnować ze swojej ulubionej trasy? . . . .	34
3.4	Analiza ruchu w godzinach szczytu . . . . .	36

3.4.1	Szczegóły analizowanego zbioru przejazdów . . . . .	38
3.4.2	Pokrycie z trasą optymalną . . . . .	38
3.4.3	Długość trasy . . . . .	39
3.4.4	Pora dnia . . . . .	40
3.4.5	Skrócenie czasu przejazdu . . . . .	41
3.4.6	Pożar mostu Łazienkowskiego i jego wpływ na ruch drogowy w Warszawie . . . . .	42
3.5	Analiza ruchu na warszawskich mostach . . . . .	43
3.5.1	Analiza różnych wariantów tras optymalnych . . . . .	43
3.5.2	Analiza korelacji . . . . .	46
3.6	Porównanie wybranych modeli klasyfikacyjnych . . . . .	48
3.6.1	Metoda k najbliższych sąsiadów . . . . .	49
3.6.2	Metoda Naive Bayes . . . . .	51
3.6.3	Drzewa decyzyjne . . . . .	51
<b>4</b>	<b>Podsumowanie</b>	<b>55</b>
4.1	Przeprowadzone prace . . . . .	55
4.2	Wnioski . . . . .	56
4.3	Kontynuacja prac . . . . .	57
	<b>Bibliografia</b>	<b>59</b>
<b>A</b>	<b>Szczegółowe wyniki</b>	<b>61</b>



# Rozdział 1

## Wstęp

### 1.1 Wprowadzenie i cel pracy

W ostatnich kilkunastu latach wśród kierowców nastąpił zdecydowany wzrost popularności systemów GPS [1]. Obecnie są one używane już nie tylko przez profesjonalnych kierowców, ale na co dzień, przez zdecydowaną większość uczestników ruchu drogowego. Jednym z powodów wzrostu popularności nawigacji jest na pewno zdecydowanie łatwiejszy dostęp do tego typu rozwiązań. Nie jest już konieczne korzystanie ze specjalnych, dedykowanych do tego celu urządzeń (choć takie wciąż są dostępne i powszechnie wykorzystywane) – wystarczy smartfon, na którym możemy zainstalować jedną z wielu, często darmowych, aplikacji mających funkcję nawigacji drogowej. Dodatkowo w dzisiejszych czasach nawigacja często działa w trybie online, co z kolei sprawia, że zazwyczaj mamy możliwość korzystania z na bieżąco aktualizowanych map oraz, co może jeszcze bardziej istotne, że podczas wyznaczania optymalnej trasy wykorzystywane są informacje o aktualnej sytuacji na drogach dostępne w czasie rzeczywistym. Rozwój tej ostatniej funkcjonalności spowodował, że obecnie kierowcy korzystają z nawigacji nawet podczas poruszania się trasami, które dobrze znają, ponieważ pozwala im ona na dostosowanie trasy przejazdu do dynamicznie zmieniającej się sytuacji na drogach.

Biorąc powyższe pod uwagę, warty zbadania wydaje się zagadnienie analizy ruchu drogowego pod kątem znalezienia pewnych wzorców czy cech charakterystycznych, które mogłyby posłużyć do stworzenia narzędzia symulującego rzeczywiste decyzje podejmowane przez kierowców przy wyborze trasy. Takie rozwiązanie mogłoby wspomagać projektowanie infrastruktury drogowej, zwłaszcza tej miejskiej. Mogłoby być także przydatne podczas analizy zmian natężenia ruchu drogowego w poszczególnych sektorach miasta po wy-

łączeniu z użycia niektórych ulic, co z kolei umożliwiłoby planowanie prac remontowych tak, aby były jak najmniej uciążliwe dla kierowców.

Celem niniejszej pracy magisterskiej jest zdefiniowanie, na podstawie analizy dostępnych danych historycznych, pewnych charakterystyk ruchu drogowego i stworzenie przy ich użyciu narzędzia pozwalającego na symulację decyzji kierowców dotyczących wyboru trasy przejazdu. W szczególności zbadane zostaną następujące kwestie:

- czy i w jakim stopniu kierowcy poruszają się po trasach optymalnych,
- wpływ informacji o bieżącym natężeniu ruchu drogowego na wybór trasy przejazdu,
- wpływ wyłączenia z użycia ważnych odcinków infrastruktury drogowej na rozkład natężenia ruchu drogowego (na przykładzie zamknięcia Mostu Łazienkowskiego),
- wybrane modele klasyfikacyjne.

## **1.2 Przegląd aktualnego stanu badań i dostępnych rozwiązań**

Problem analizy ruchu drogowego jest stosunkowo złożonym zagadnieniem. Wynika to przede wszystkim z faktu, że na zachowanie i decyzje kierowców na drodze ma wpływ bardzo wiele czynników, które często trudno jest precyzyjnie zdefiniować, a jeszcze trudniej dokładnie zmierzyć. Na przestrzeni lat powstało wiele opracowań dotyczących tego tematu, a spektrum aspektów, które są w nich poruszane jest bardzo różnorodne. W dalszej części przedstawionych zostało kilka pozycji, które są związane z tematem niniejszej pracy.

### **1.2.1 Paradoks Braessa**

Jak już wspomniano wcześniej jedną z motywacji dla powstania tej pracy było stworzenie narzędzia, które pomagałoby podczas planowania wyłączenia pewnych odcinków dróg tak, aby było to jak najmniej uciążliwe dla kierowców. W tym kontekście warto wspomnieć o paradoksie Braessa [2, 3], który opisuje sytuację, kiedy to otwarcie nowego odcinka drogi, w sytuacji, gdy drogi w okolicy są przeciążone z powodu nadmiernej liczby samochodów, powoduje, że wydłużają się czasy przejazdów wszystkich pojazdów w analizowanym obszarze dróg. Takie zjawisko może wydawać się zupełnie sprzeczne

z intuicją, ale znane są przypadki, gdy w rzeczywistym ruchu drogowym obserwowano efekty opisane w paradoksie Braessa. Przykładem może tu być czasowe zamknięcie 42. ulicy w Nowym Jorku w 1990 roku [4]. Potwierdza to tylko jak złożoną kwestią jest analiza ruchu drogowego i pokazuje jak ciężko może być zaplanować najlepszy rozkład ulic.

### 1.2.2 Informacja o aktualnym natężeniu ruchu drogowego kiedyś a dziś

W dzisiejszych czasach, dzięki smartfonom z funkcją GPS i dostępem do Internetu oraz wielu, często darmowym, aplikacjom z funkcją nawigacji, kierowcy mają bardzo łatwy dostęp do dokładnych danych przedstawiających sytuację na drogach w czasie rzeczywistym. Jednak jeszcze kilka, a już na pewno kilkanaście lat temu, sytuacja wyglądała zupełnie inaczej. W czasach sprzed upowszechnienia się systemu GPS zdecydowanie większe znaczenie miały na przykład serwisy informacyjne nadawane w lokalnych stacjach radiowych, które przekazywały kierowcom raporty o aktualnej sytuacji na drogach w regionie.

Badania naukowe dotyczące analizy zachowań kierowców i tego z jakich informacji korzystają planując podróż prowadzone są od wielu lat. Dotyczą one tego samego zagadnienia, ale na przestrzeni lat, wraz ze zmieniającymi się metodami dostępu do danych na temat ruchu drogowego, zmieniała się zarówno metodologia badań, jak i kwestie, które są poddawane szczegółowej analizie.

Upowszechnienie się systemów GPS, które dostarczają aktualnych informacji o ruchu drogowym przy wykorzystaniu sieci Internet, nie tylko sprawiło, że zmieniło się podstawowe źródło informacji o bieżącej sytuacji na drogach wykorzystywane przez kierowców, ale, co może być jeszcze istotniejsze z badawczego punktu widzenia, dostarczyło dokładnych informacji na temat zachowania na drodze i decyzji podejmowanych przy wyborze trasy. Wynika to z faktu, że systemy nawigacji, aby móc dostarczać wiarygodnych informacji o aktualnej sytuacji na drogach pozyskują i przetwarzają informacje od użytkowników. W ten sposób powstają duże, bardzo ciekawe z analitycznego punktu widzenia, wolumeny danych zawierające dokładne informacje o położeniu użytkowników zmieniającym się w czasie.

Nowy sposób pozyskiwania danych zrewolucjonizował w pewnym sensie podejście do badań dotyczących zagadnień związanych z zachowaniem kierowców i ich nawykami. Wcześniej często podstawą do przeprowadzanych analiz były ankiety wypełniane przez kierowców. Taka metodologia ma jednak wiele niedoskonałości. Rozmiar badanej próby jest zazwyczaj stosunkowo

niewielki, gdyż zachęcenie dużej liczby osób do wzięcia udziału w badaniu jest czasochłonne, kosztowne i niezbyt wygodne. Kolejną kwestią jest, że pozyskiwane w ten sposób informacje nie są zbyt precyzyjne, gdyż respondenci mogą w różny sposób interpretować pytania czy skalę odpowiedzi. Dodatkowo nie ma gwarancji, że odpowiadają zgodnie z prawdą.

Wszystkie powyższe problemy nie występują, gdy badania są prowadzone na podstawie danych zebranych za pomocą systemów GPS. W takim przypadku dane zbieramy niejako „przy okazji” i są one wiarygodne, a także bardzo precyzyjne.

### **1.2.3 Analiza i inżynieria ruchu na podstawie ankiet**

Tak jak uprzednio wspomniano, przed upowszechnieniem się systemów nawigacji GPS, wiele prac dotyczących czynników wpływających na planowanie podróży opierało się na danych zebranych od respondentów za pomocą ankiet.

Interesujące zagadnienie, blisko związane z tematem niniejszej pracy, poruszone zostało w artykule [5], gdzie autorzy badali wpływ informacji o aktualnej sytuacji na drogach dostarczanych w czasie rzeczywistym na decyzje kierowców co do zmiany trasy, bądź zmiany godziny wyjazdu. Praca dotyczyła porannych godzin szczytu komunikacyjnego na obszarze centrum Chicago, a analizowanym źródłem informacji o aktualnym natężeniu ruchu drogowego dla kierowców były drogowe serwisy informacyjne nadawane przez lokalne radiostacje. W kwestionariuszu ankietowani odpowiadali na pytania dotyczące tego czy radiowe raporty o sytuacji na drogach wpływają na ich decyzje odnośnie wyboru trasy i godziny wyjazdu oraz oceniali ogólny poziom satysfakcji z działania systemu informacji drogowych, a także wskazywali jakie ulepszenia w jego działaniu byłyby pożądane.

Kolejne badanie [6], którego współautorem również jest profesor Khatkhat, dotyczyło wpływu niekorzystnych warunków pogodowych na skłonność kierowców do zmiany swoich przyzwyczajzeń. Autorzy analizują jakie czynniki wpływają na możliwość zmiany środka transportu, trasy czy godziny wyjazdu. Źródłem danych również była ankieta, tym razem przeprowadzona wśród pracowników sektora publicznego i prywatnego, a także osób samozatrudnionych pracujących w Brukseli. Przygotowany raport dostarcza informacji jaki wpływ na decyzje kierowców, w powyższym kontekście, mają czynniki socjoekonomiczne. Twórcy, w swojej pracy, podkreślają, że temat wpływu warunków atmosferycznych na zachowania kierowców jest zagadnieniem, na temat którego ciężko jest znaleźć inne opracowania naukowe. Jednym ze sformułowanych wniosków jest to, iż ankietowani z powodu niekorzystnej pogody zdecydowanie chętniej zmieniają godzinę wyjazdu niż trasę,

którą poruszają się zazwyczaj, czy swój typowy środek transportu.

Schemat przeprowadzania badań naukowych nad zachowaniami kierowców na podstawie ankiet jest w dużej mierze jednolity. Różnice zazwyczaj polegają na szczegółach przyjętej metodologii i aspektach, na których skupia się badanie.

Na Tajwanie naukowcy skoncentrowali się tylko na analizie decyzji kierowców podróżujących jedną z dróg ekspresowych [8]. Kolejnym elementem charakterystycznym dla tego badania było zbieranie danych w trzech różnych okresach: w dni robocze, w czasie weekendu i w trakcie trwania wiosennych ferii. Podobnie jak we wcześniej wspomnianych pracach, obserwowany był wpływ informacji o natężeniu ruchu drogowego dostarczanych w czasie rzeczywistym na skłonność kierowców do zmiany trasy. Aktualne informacje o sytuacji na drogach były przekazywane za pomocą raportów w rozgłoszeniach radiowych oraz przy użyciu specjalnych znaków z dynamiczną treścią. Dodatkowo autorzy wyróżnili pięć rodzajów informacji o bieżącej sytuacji na drodze (od ogólnych i lakonicznych, po dokładne zalecenia co do zmiany trasy na nową) i sprawdzali jak każdy z tych rodzajów wpływa na skłonność kierowców do zmiany trasy. Jedną z konkluzji była zależność, iż im dokładniejsze informacje na temat sytuacji na drogach, tym kierowcy będą bardziej skorzy do zmiany trasy. Co ciekawe, autorzy zauważają, że zależność ta nie jest jednak bardzo wyraźna, dużo mniej charakterystyczna niż w innych badaniach, na które się powołują. Przyczyn takiego stanu rzeczy dopatrują się w fakcie, że w czasach przeprowadzania badań informacje o natężeniu ruchu drogowego podawane w czasie rzeczywistym były dla kierowców na Tajwanie czymś nowym i dlatego mogli podchodzić do nich z większą rezerwą. To pokazuje tylko jak złożonym zagadnieniem może być modelowanie ruchu i jak różne czynniki należy brać pod uwagę podczas prac nad tym tematem.

#### 1.2.4 Modelowanie ruchu drogowego

W artykule [11] przedstawiony został opis stworzonego modelu symulującego decyzje kierowców korzystających z zaawansowanego systemu informacji podróżnych (ang. *Advanced Traveler Information System*). Autorzy rozważają wpływ niechęci (ang. *inertia*) oraz uległości (ang. *compliance*) do kierowania się informacjami z ATIS. Niechęć jest definiowana jako poruszanie się po zdefiniowanej trasie domyślnej, natomiast uległość jako podatność na zmianę trasy domyślnej pod wpływem informacji o bieżącej sytuacji na drogach dostarczanej w czasie rzeczywistym. Opracowane narzędzie może znajdować zastosowanie w przewidywaniu wpływu systemów ATIS na płynność ruchu. Twórcy zaznaczają jednak, że wyniki symulacji teoretycznych powinny zostać skonfrontowane z doświadczalnymi obserwacjami ze świata

rzeczywistego. Naukowcy podkreślają również, że stworzony przez nich model był opracowywany pod kątem jak najlepszego odzwierciedlenia ruchu kierowców dojeżdżających do pracy w czasie porannego szczytu komunikacyjnego. Charakterystyka takich przejazdów może być znacząco różna od podróży wakacyjnych czy przejazdów odbywających się w dni wolne od pracy. Jednym z wniosków płynących z zastosowania modelu była obserwacja, że im większa oszczędność czasu wynika ze zmiany trasy, tym kierowcy są bardziej skory do takiej zmiany.

Trochę inna kwestia, związana z modelowaniem ruchu drogowego, została poruszona w artykule [12]. Przedstawia on zaawansowany system routingu, uwzględniający w czasie rzeczywistym informacje o aktualnej sytuacji na drogach oraz indywidualne preferencje kierowcy. Autorzy podkreślają, że wyznaczanie trasy optymalnej jest wyjątkowo złożonym zagadnieniem i jak pokazują odpowiedzi ankietowanych, wcale nie sprowadza się tylko do wyboru najszybszej bądź też najkrótszej trasy. W zaproponowanym modelu zostały uwzględnione takie czynniki charakteryzujące trasę jak:

- długość trasy,
- czas przejazdu,
- natężenie ruchu,
- opłaty drogowe,
- poziom trudności trasy (szerokość jezdni, liczba pasów, itp.),
- krajobraz.

Kierowca korzystający z systemu podaje swoje preferencje odnośnie powyższych atrybutów trasy. Opracowany model opiera się na sieci neuronowej połączonej z logiką rozmytą, co sprawia, że system, wraz z upływem czasu, „uczy się” na podstawie analizy rzeczywistych tras wybieranych przez kierowcę i w ten sposób dostosowuje swoje działanie do jego indywidualnych potrzeb. Zaproponowana metodologia jest jednym z przykładów inteligentnych systemów nawigacji.

### **1.2.5 Informacje w czasie rzeczywistym a płynność ruchu**

Badając zagadnienia inżynierii ruchu drogowego wydaje się, że informacja o bieżącej sytuacji na drogach może pozytywnie wpływać na płynność ruchu

[5]. Jednak według badań, które analizują wpływ systemów GPS dostarczających informacje o utrudnieniach na drogach w czasie rzeczywistym, wcale nie musi tak być. Naukowcy z Uniwersytetu Kalifornijskiego twierdzą [7], że popularność smartfonów i aplikacji z usługą nawigacji sprawiła, że kierowcy często poruszają się po trasach, które nie są zgodne z intencjami planistów infrastruktury drogowej. Obserwowany jest nadmierny wzrost natężenia ruchu na drogach, które są do tego nieprzystosowane (drogi lokalne, uliczki osiedlowe, itp.). Z przeprowadzonych badań i obserwacji wynika, że stosowanie się przez kierowców na masową skalę do wskazówek systemów nawigacji uwzględniających sytuację na drogach w czasie rzeczywistym, może lokalnie powodować jeszcze większe utrudnienia w ruchu niż miałyoby to miejsce, gdyby kierowcy z nich nie korzystali. Autorzy podkreślają jednak, że niezbadana pozostaje kwestia wpływu na płynność ruchu, jeśli będziemy analizowali sytuację na większym obszarze.

### 1.2.6 Modele predykcji ruchu a zachowanie kierowców

Jednym z powszechnie przyjmowanych założeń w zagadnieniach związanych z inżynierią ruchu drogowego jest to, że kierowcy poruszają się po trasach optymalnych. Na tej tezie oparte jest działanie większości modeli służących do przewidywania natężenia ruchu drogowego. Takie założenie może wydawać się także zgodne z intuicją, bo przecież kierowcy racjonalnie podejmujący decyzję powinni wybierać trasy optymalne.

Jednak z drugiej strony, kwestia wyboru trasy przejazdu jest bardzo złożona i ma na nią wpływ wiele czynników, w dużej mierze indywidualnych dla każdego kierowcy. Dodatkowo ciężko może być obiektywnie stwierdzić, która trasa jest optymalna, ponieważ może to zależeć od zmieniających się dynamicznie warunków, takich jak na przykład natężenie ruchu, prowadzone roboty drogowe czy utrudnienia związane z wypadkami komunikacyjnymi.

Jak pokazują badania [9] kierowcy wcale nie jeżdżą zgodnie z trasami optymalnymi wyznaczonymi według przyjętego kryterium. Podobne wnioski, a mianowicie, że modele powszechnie używane do predykcji ruchu odbiegają od obserwowanego w rzeczywistości ruchu drogowego, są formułowane w artykułach profesora Khattaka [6].

Biorąc powyższe pod uwagę, w niniejszej pracy przeanalizowane zostało, na podstawie dostępnych danych, również to, czy rzeczywiste przejazdy pokrywają się z wyznaczonymi trasami optymalnymi.

### 1.2.7 Samochodowe systemy GPS a modelowanie ruchu

W ostatnich latach pojawiała się całkowicie nowa możliwość pozyskiwania danych, które mogą być następnie wykorzystywane w analizach ruchu drogowego. W prywatnych samochodach na coraz większą skalę zaczęły pojawiać się systemy GPS posiadające jednocześnie łączność z Internetem. Przykład takiego rozwiązania obsługiwanego przez firmę OCTOTElematics oraz jego wykorzystanie w zagadnieniach związanych z analizą i modelowaniem ruchu opisano w [13].

Autorzy przedstawiają dwa sposoby predykcji prędkości w krótkiej perspektywie czasowej na poszczególnych odcinkach dróg. W tym celu wykorzystane zostały dane pozyskane z 600 000 dedykowanych urządzeń (wyposażonych m. in. w odbiorniki GPS i moduły GPRS) zamontowanych w samochodach prywatnych użytkowników w porozumieniu z firmami ubezpieczeniowymi. W artykule podkreślono, że dzięki powyższemu rozwiązaniu możliwe jest gromadzenie w efektywny sposób dużej ilości dokładnych danych opisujących ruch drogowy. Takie podejście ma duży potencjał i jest tańsze od stosowanych uprzednio w systemach ATIS metod pozyskiwania danych, takich jak: monitoring video, pętle indukcyjne, czujniki radarowe, itp., gdyż nie wymaga żadnej ingerencji w infrastrukturę drogową.

W pracy porównano modelowanie prędkości, z jaką poruszają się samochody na autostradowej obwodnicy Rzymu, przy wykorzystaniu techniki dopasowania do wzorca (ang. *pattern matching*) oraz sieci neuronowych. Wyniki przeprowadzonych badań pokazują, że obie metody mogą znaleźć zastosowanie w różnych rozwiązaniach typu ATIS, jednak kluczowe znaczenie dla dokładności uzyskiwanych predykcji ma wielkość próbki danych, a dokładniej stosunek liczby monitorowanych pojazdów do ich ogólnej liczby powinien być możliwie wysoki.

### 1.2.8 Efektywne modelowanie ruchu drogowego w czasie rzeczywistym

Możliwość pozyskiwania w czasie rzeczywistym informacji o aktualnej sytuacji na drogach z systemów GPS zamontowanych w samochodach na masową skalę spowodowała, że dostępność danych przestała być problemem. W związku z tym, w ostatnich latach w badaniach z dziedziny inżynierii ruchu więcej uwagi zostało skierowane na aspekt użycia dostępnych danych w celu poprawy efektywnego wykorzystania infrastruktury drogowej.

Badania w tym zakresie były prowadzone między innymi w oddziale badawczym firmy IBM. Autorzy raportu [14] przedstawiają opracowany czasowo-



przestrzenny model predykcji ruchu drogowego. Podstawowym założeniem zaproponowanego rozwiązania była wysoka częstotliwość aktualizacji danych modelu (5 minut, gdzie typowo w innych rozwiązaniach przyjmowane jest 15 minut), oraz uniwersalność w takich aspektach jak: pora dnia, dzień tygodnia, typ drogi, a także predykcja dla wielu punktów odniesienia w przyszłości. Ta ostatnia cecha jest szczególnie istotna z punktu widzenia użyteczności przygotowanego rozwiązania dla kierowców poruszających się po dłuższych trasach. Dodatkowo, przygotowany model nie mógł cechować się zbyt dużą złożonością obliczeniową, gdyż wymogiem było, aby wyniki były dostarczane w czasie rzeczywistym.

Przedstawione efekty prac są obiecujące, gdyż sprawdzenie działania na testowym zbiorze danych przyniosło bardzo dobre rezultaty – model pracował wydajnie i z dużą dokładnością. Jako kontynuacja prac zapowiadane jest rozszerzenie przygotowanego rozwiązania o uwzględnianie przy predykcji ruchu drogowego informacji o pogodzie, wypadkach i robotach drogowych.



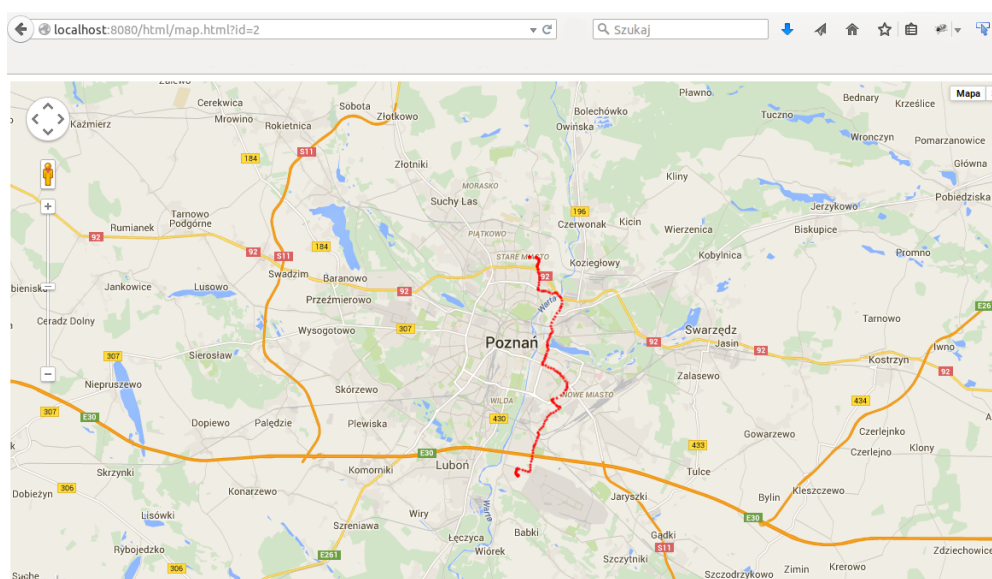
## Rozdział 2

# Stworzenie środowiska analitycznego

Bazą do rozpoczęcia pracy z dostępnymi danymi był wybór i stworzenie odpowiednich narzędzi, a mówiąc jeszcze szerzej, kompleksowego środowiska analitycznego. Przyjęte zostało założenie, że nie będą wykorzystywane rozwiązania, które oferowane są jako usługi. Przyjęcie takiej koncepcji miało na celu uniknięcie sytuacji braku kontroli nad wykorzystywanymi rozwiązaniami. Wybór i stworzenie przydatnych narzędzi okazał się procesem trudnym i czasochłonnym, podczas którego wielokrotnie podejmowane decyzje okazywały się nietrafnymi, a testowane rozwiązania nieadekwatnymi do potrzeb. Finalnie udało się jednak stworzyć środowisko, które posłużyło do kompleksowej analizy danych i badania czynników determinujących zachowania kierowców.

### 2.1 Przegląd wykorzystywanych narzędzi

Wykorzystywane w pracy dane na temat ruchu drogowego pochodzą z aplikacji Yanosik [15] i zostały udostępnione do badań w formie plików tekstowych CSV. Wybrane narzędzia i technologie miały umożliwić ich przetwarzanie i analizę. Podczas kolejnych etapów prac, uwaga była koncentrowana na innych aspektach badań, od analizy pojedynczych przejazdów wykreślanych na mapach, do wsadowego przetwarzania wcześniej wyodrębnionych grup przejazdów. Poniżej przedstawiono krótki opis wykorzystywanych technologii i ich zastosowanie w pracach na tym temacie.



Rys. 1: Czerwonymi punktami wykreślona trasa jednego z analizowanych przejazdów

## System zarządzania bazą danych - MySQL

W celu umożliwienia łatwego przetwarzania dużej ilości danych, dostarczonych w formie plików tekstowych, zaimportowano je do relacyjnej bazy danych. Jako system zarządzania bazą danych wybrano MySQL [16], który jest popularnym rozwiązaniem udostępnianym na zasadach wolnego oprogramowania.

## Wizualizacja tras przejazdów

Analiza danych została rozpoczęta od badania pojedynczych przejazdów. W tym celu zostało stworzone proste narzędzie oparte o technologie internetowe, które okazało się być przydatne właściwie podczas wszystkich kolejnych etapów prac. Umożliwia ono wykreślanie na mapie tras przejazdów pobranych z bazy danych i wyświetlenie ich w przeglądarce internetowej (Rys. 1). Wykorzystane w tym celu technologie to HTML, JavaScript, PHP, Google Maps API oraz serwer WWW Apache.

## QGIS

Na etapie wyboru wykorzystywanych narzędzi, sprawdzone zostały możliwości programu QGIS [17] oraz dostępna do niego wtyczka RoadGraph. Potencjalnie przydatną funkcją była możliwość wyznaczania tras optymalnych

między zadanymi punktami. Po zapoznaniu się z powyższym oprogramowaniem okazało się jednak, że planowane zadania prawdopodobnie łatwiej będzie wykonać wybierając inne rozwiązania, dlatego pomysł wykorzystania QGIS-a został zarzucony.

## GraphHopper

Alternatywą dla programu QGIS okazała się biblioteka GraphHopper [18]. Jest ona napisana w języku Java i udostępniana jako wolne oprogramowanie. Umożliwia łatwe wyznaczanie tras pomiędzy zadanymi punktami przy wykorzystaniu informacji geograficznych dostępnych w ramach projektu OpenStreetMap. W oparciu o tę bibliotekę stworzona została aplikacja w języku Java, która posłużyła jako główne narzędzie do analizy zagadnień poruszanych w pracy. Zaletą takiego podejścia była dużo większa elastyczność rozwiązania i możliwość dopasowania do indywidualnych wymagań, niż miałyby to miejsce przy użyciu programu QGIS.

## R

W celu analizy danych, zarówno tych surowych, pobieranych bezpośrednio z bazy danych, jak i tych już przetworzonych, powstałych jako wyniki działania stworzonej aplikacji, używane były skrypty napisane w języku R [19]. Środowisko R, a w szczególności pakiet Caret [20], zostało również wykorzystane do opracowania modeli klasyfikacyjnych.

## 2.2 Środowisko analityczne

Jako główne narzędzie do analizy danych posłużyła aplikacja napisana w języku Java i wykorzystująca bibliotekę GraphHopper. Podczas implementacji pojawiło się kilka bardziej złożonych zagadnień, dla których wybrane rozwiązania opisano poniżej.

### 2.2.1 Definicje wykorzystywanych pojęć

W celu usystematyzowania pewnych pojęć, które będą wykorzystywane w dalszej części pracy przyjęto poniższe definicje i oznaczenia:

- $S = \{s_1, s_2, s_3, \dots, s_n\}$  – zbiór przejazdów  $s_i$ ,
- $s_i = \{\mathbf{x}_1(t_1), \mathbf{x}_2(t_2), \mathbf{x}_3(t_3), \dots, \mathbf{x}_n(t_n)\}$  – pojedynczy przejazd opisany jako zbiór punktów  $\mathbf{x}_j(t_j)$  definiujących położenie pojazdu w chwili  $t_j$ ,

- $G = (U, E, W)$  – graf reprezentujący sieć ulic, stworzony na podstawie danych z OpenStreetMap,
- $U = \{u_1, u_2, u_3, \dots, u_n\}$  – zbiór wierzchołków  $u_i$  grafu  $G$  reprezentujący skrzyżowania,
- $E = \{e_1, e_2, e_3, \dots, e_n\}$  – zbiór krawędzi  $e_i$  grafu  $G$ ,
- $W = \{w_1, w_2, w_3, \dots, w_n\}$  – zbiór wag krawędzi  $w_i$  grafu  $G$ ,
- $e_i = (u_j, u_k)$  – krawędź grafu  $G$  reprezentująca odcinek drogi pomiędzy dwoma skrzyżowaniami  $u_j$  i  $u_k$ ,
- $P_i = \phi_G(\mathbf{x}_i, \mathbf{x}_j) = (u_k, u_l, u_m, \dots)$  – najkrótsza ścieżka znaleziona pomiędzy punktami  $\mathbf{x}_i$  i  $\mathbf{x}_j$  reprezentowana jako zbiór kolejnych wierzchołków grafu  $G$ ,
- $P_{B_i} = \phi_{G_B}(\mathbf{x}_i, \mathbf{x}_j)$  – najszybsza trasa znaleziona pomiędzy punktami  $\mathbf{x}_i$  i  $\mathbf{x}_j$  bez uwzględnienia aktualnego natężenia ruchu drogowego,
- $P_{T_i} = \phi_{G_T}(\mathbf{x}_i, \mathbf{x}_j)$  – najszybsza trasa znaleziona pomiędzy punktami  $\mathbf{x}_i$  i  $\mathbf{x}_j$  przy uwzględnieniu aktualnego natężenia ruchu drogowego,
- $R_j = \psi_G(s_j) = (e_k, e_l, e_m)$  – projekcja śladu przejazdu  $s_j$  na graf  $G$ ,
- $l_i = (\mathbf{x}_j(t_j), \mathbf{x}_k(t_k))$  – odcinek pomiędzy punktami  $\mathbf{x}_j(t_j)$  i  $\mathbf{x}_k(t_k)$  reprezentujący fragment przejazdu,
- $\tilde{l}_j = (\mathbf{x}_m(t_m), \mathbf{x}_n(t_n))$  – odcinek pomiędzy punktami  $(\mathbf{x}_m(t_m), \mathbf{x}_n(t_n))$  reprezentujący fragment przejazdu, który pokrywa się z fragmentem wyznaczonej trasy optymalnej,
- $C = \frac{\sum_{j=1}^l |\tilde{l}_j|}{\sum_{i=1}^k |l_i|}$  - współczynnik pokrycia trasy przejazdu z trasą optymalną.

Powyższe oznaczenia i wzory będą wykorzystywane w dalszej części pracy, gdzie zostaną szerzej omówione.

### 2.2.2 Wyznaczanie tras optymalnych

Jednym z powodów, dla którego aplikacja służąca do analizy danych ruchowych została opracowana w oparciu o wykorzystanie biblioteki Graph-Hopper, była łatwość wyznaczania trasy optymalnej między zadanymi punktami, jaką daje to rozwiązanie. Przy takim podejściu, poszukiwanie trasy

optymalnej sprowadza się do skonstruowania zapytania, gdzie należy określić podstawowe parametry, takie jak:

- współrzędne początku i końca trasy,
- typ trasy – najkrótsza lub najszybsza,
- środek lokomocji,
- algorytm poszukiwania najkrótszej ścieżki.

W przypadku zagadnień poruszanych podczas analiz właściwie wszystkie z powyższych atrybutów były stałe, oczywiście za wyjątkiem punktu początkowego i końcowego przejazdu. W większości przeprowadzonych badań rozważana była trasa najszybsza, a środkiem transportu, brany pod uwagę, był samochód. Natomiast metodą używaną do poszukiwania optymalnej trasy był algorytm Dijkstry w wariacie dwukierunkowym.

### Algorytm Dijkstry

Algorytm Dijkstry [21] został opracowany przez holenderskiego informatyka Edsgara Dijkstrę i może być on wykorzystywany do poszukiwania najkrótszej ścieżki pomiędzy dwoma zadanymi wierzchołkami w grafie o nieujemnych wagach krawędzi. Zasada działania algorytmu przedstawia się następująco:

1. Ustalane są odległości od wierzchołka początkowego  $u_0$  dla wszystkich wierzchołków  $u$  w grafie jako równe  $\infty$ , a odległość  $d[u_0]$  jako równa 0.
2. Tworzona jest kolejka priorytetowa składająca się ze wszystkich wierzchołków, gdzie priorytetem jest aktualnie wyznaczona odległość danego wierzchołka od wierzchołka  $u_0$ .
3. Dopóki w kolejce znajdują się elementy wybierany jest wierzchołek  $u_i$  najbliższy wierzchołkowi  $u_0$ , który nie został jeszcze rozważony i dla wszystkich sąsiadów  $u$  wybranego wierzchołka  $u_i$  dokonywane jest sprawdzenie czy odległość  $d[u_i] + w(u_i, u) < d[u]$ , jeśli jest to aktualizowana jest  $d[u]$ .

Algorytm Dijkstry znajduje najkrótsze ścieżki w grafie z wybranego wierzchołka  $u_0$  do wszystkich pozostałych wierzchołków w grafie. W przypadku, gdy szukana jest najkrótsza ścieżka między dwoma zadanymi wierzchołkami, działanie algorytmu jest przerywane w momencie znalezienia najkrótszej ścieżki z  $u_0$  do drugiego zadanego wierzchołka.

Jedną z technik przyspieszenia działania algorytmu Dijkstry w praktyce jest jego wersja w formie dwukierunkowej. Modyfikacja ta polega na równoczesnym przetwarzaniu wierzchołków zaczynając zarówno od wierzchołka początkowego, jak i od wierzchołka końcowego na odwróconym grafie.

### Algorytm A\*

Drugim, dostępnym do wyboru, algorytmem poszukiwania najkrótszej ścieżki, jaki został zaimplementowany w bibliotece GraphHopper, jest algorytm A\* [22]. Na początkowym etapie prac porównane zostało działanie algorytmu Dijkstry i algorytmu A\*. Otrzymane na próbie testowej rezultaty pokazały jednak, że wyznaczone przy użyciu różnych algorytmów trasy optymalne były identyczne. Jak wspomniano wcześniej, do dalszego wykorzystania wybrany został algorytm Dijkstry, w związku z czym, algorytm A\* nie będzie szerzej opisywany.

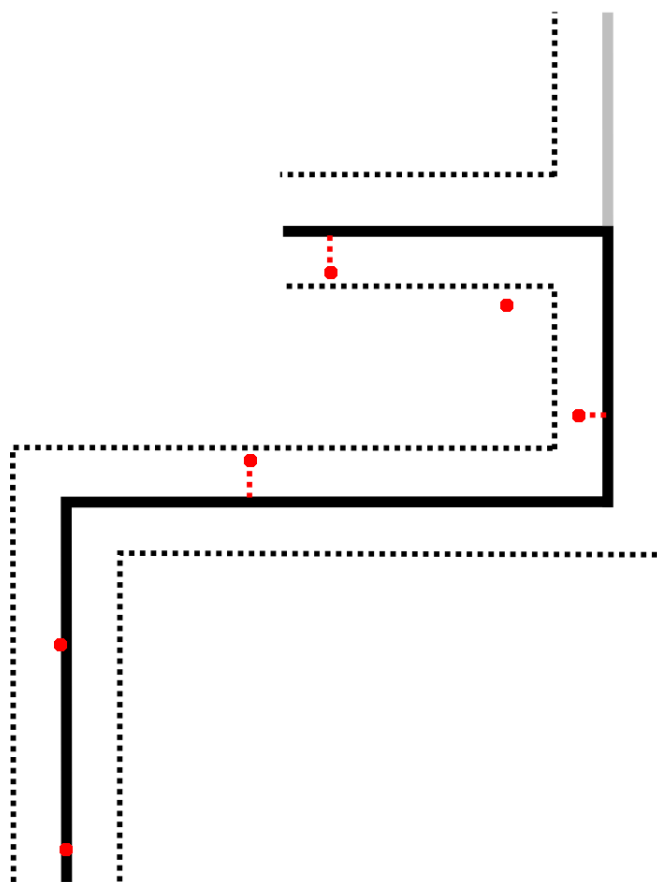
### 2.2.3 Wyznaczanie części wspólnych tras

Jedną z pierwszych badanych kwestii było sprawdzenie w jakim stopniu rzeczywiste trasy przejazdów pokrywają się z trasami optymalnymi wyznaczonymi na różne sposoby. Trasa przejazdu zapisana poprzez aplikację Yanosik jest przedstawiona w postaci zbioru kolejnych punktów opisujących położenie w danej chwili. Przyjęte zostało dla niej oznaczenie  $s_i = \{\mathbf{x}_1(t_1), \mathbf{x}_2(t_2), \mathbf{x}_3(t_3), \dots, \mathbf{x}_n(t_n)\}$ . Natomiast trasa wyznaczona przy użyciu biblioteki GraphHopper ma postać zbioru odcinków  $o_i = \{e_1, e_2, e_3, \dots, e_k\}$ . Taka różnica w reprezentacji tras sprawiała, że określenie w jakiej części są one tożsame było nietrywialne. Należało zdefiniować funkcję  $\psi_G(s_i)$ , która pozwalałaby na projekcję śladu  $s_i$  na graf  $G$ . Szczegółowa implementacja została przedstawiona poniżej, a na rysunku (Rys. 2) przedstawiono w sposób poglądowy wykorzystywany mechanizm rzutowania punktów.

Dla każdego punktu opisującego trasę rzeczywistą:

1. Wyznaczane są jego współrzędne  $x$  i  $y$  w projekcji sinusoidalnej [23].
2. Obliczana jest odległość od punktu poprzedniego.
3. Iteracja po kolejnych odcinkach wyznaczonej trasy optymalnej – dla każdego odcinka:
  - (a) Wyznaczane są współrzędne punktu początkowego i końcowego odcinka analogicznie jak dla punktu trasy rzeczywistej.





Rys. 2: Poglądowe przedstawienie rzutowania punktów opisujących rzeczywistą trasę przejazdu na odcinki tworzące siatkę ulic w rozwiązaniu OpenStreetMap. Na rysunku zastosowano następujące oznaczenia:

- linią ciągłą w kolorze czarnym zaznaczono odcinki siatki ulic OpenStreetMap zaklasyfikowane jako te, po których odbywał się przejazd,
- linią ciągłą w kolorze szarym zaznaczono pozostałe odcinki siatki ulic OpenStreetMap,
- kolorem czerwonym zaznaczono punkty trasy rzeczywistej wraz z ich rzutami prostokątnymi na odcinki siatki ulic OpenStreetMap,
- linią przerywaną w kolorze czarnym zaznaczono obszar (zgodnie z zadaniem progiem dokładności), w którym musi znaleźć się punkt, aby mógł być on klasyfikowany jak znajdujący się na danym odcinku siatki ulic OpenStreetMap.

- (b) Na podstawie otrzymanych wcześniej punktów wyznaczone jest równanie prostej w postaci ogólnej.
  - (c) Wyznaczana jest prosta prostopadła do zadanej prostej przechodząca przez punkt trasy rzeczywistej.
  - (d) Powstały układ równań jest rozwiązywany metodą wyznaczników.
  - (e) Wykonywane jest sprawdzenie, czy znaleziony punkt przecięcia prostych leży na zadanym odcinku trasy optymalnej.
  - (f) Jeśli tak, to obliczana jest odległość punktu trasy rzeczywistej od prostej zawierającej odcinek trasy optymalnej.
  - (g) Wykonywane jest sprawdzenie, czy obliczona odległość jest mniejsza od odległości obliczonej w poprzednich krokach i jeśli tak, to aktualizowana jest minimalna wyznaczona odległość.
4. Wykonywane jest sprawdzenie, czy znaleziona minimalna odległość jest mniejsza od zadanego progu ( $\sim 100\text{m}$ ).
  5. Jeśli tak, to jest uznawane, że dana część trasy rzeczywistej pokrywa się z trasą optymalną.
  6. Długość danego odcinka trasy rzeczywistej jest sumowana z wcześniejszymi odcinkami, które też pokrywają się z trasą optymalną.

Na koniec obliczane jest w jakim stopniu trasa rzeczywistego przejazdu pokrywa się z wyznaczoną trasą optymalną, co można zapisać za pomocą wzoru:

$$C = \frac{\sum_{j=1}^L |\tilde{l}_j|}{\sum_{i=1}^K |l_i|}, \quad (2.1)$$

gdzie:

- $|l_i| = |(\mathbf{x}_j(t_j), \mathbf{x}_k(t_k))|$  – długość odcinka reprezentującego fragment przejazdu pomiędzy punktami  $\mathbf{x}_j(t_j)$  i  $\mathbf{x}_k(t_k)$ ,
- $|\tilde{l}_j| = |(\mathbf{x}_m(t_m), \mathbf{x}_n(t_n))|$  – długość odcinka reprezentującego fragment przejazdu pomiędzy punktami  $(\mathbf{x}_m(t_m), \mathbf{x}_n(t_n))$ , który pokrywa się z fragmentem wyznaczonej trasy optymalnej,
- $K$  – liczba odcinków składająca się na dany przejazd  $s_i$ ,
- $L$  – liczba odcinków pokrywających się z wyznaczoną trasą optymalną dla danego przejazdu  $s_i$ .

### 2.2.4 Modelowanie natężenia ruchu

Do wyznaczania tras optymalnych biblioteka GraphHopper wykorzystuje dane z OpenStreetMap. Drogi w projekcie OpenStreetMap reprezentowane są jako zbiór krawędzi  $E$ . Do każdej krawędzi  $e_i$  przypisana jest odpowiednia waga  $w_i$ , która jest używana przez algorytmy wyznaczania trasy optymalnej. Jednym z zagadnień poruszanych w niniejszej pracy jest wyznaczanie tras optymalnych przy uwzględnieniu aktualnego natężenia ruchu drogowego. Do modelowania aktualnej sytuacji na drogach posłużyły udostępnione dane pochodzące z aplikacji Yanosik.

Natężenie ruchu na drogach modelowane było w piętnastominutowych oknach czasowych. Z każdym z punktów pomiarowych opisujących przejazdy rzeczywiste związana jest aktualna prędkość pojazdu w danej chwili. Ta wartość posłużyła do odwzorowania aktualnego stanu na drogach w danej chwili. Dla każdego z punktów pomiarowych w wybranym oknie czasowym została znaleziona krawędź reprezentująca odcinek drogi, po którym pojazd poruszał się w danym momencie. Zastosowany sposób modyfikacji prędkości przypisanej do danego odcinka drogi można wyrazić następującym wzorem:

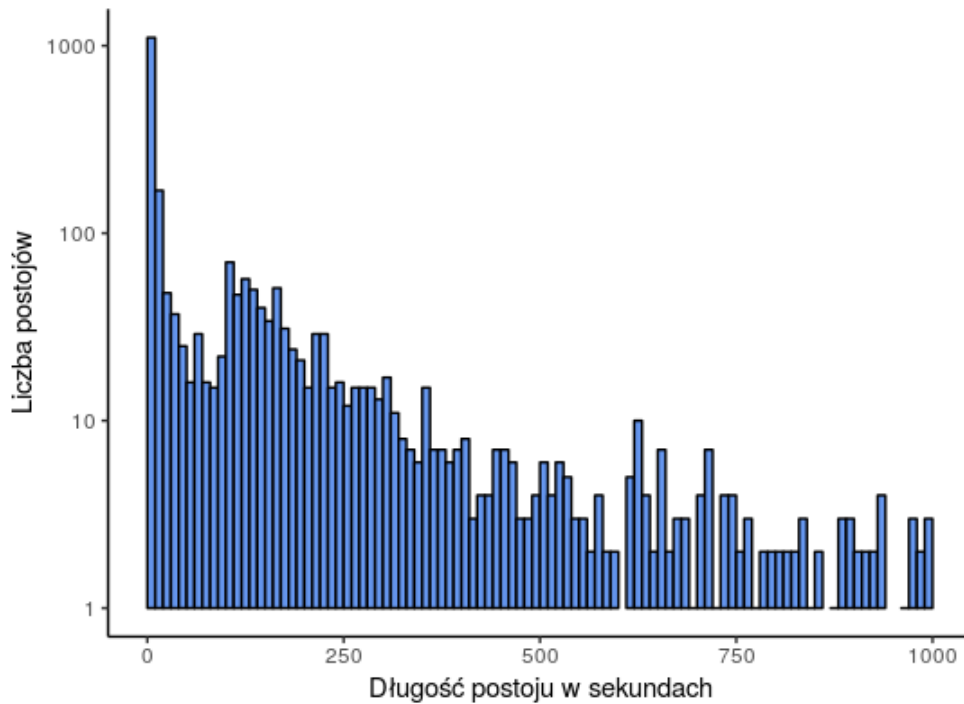
$$v = \frac{2v_{osm} + \sum_{i=1}^k v_i}{k + 2}, \quad (2.2)$$

gdzie:

- $v_{osm}$  – prędkość przypisana do danego odcinka drogi według danych z OpenStreetMap,
- $v_i$  – rzeczywista prędkość poruszania się pojazdu przypisana do danego punktu pomiarowego,
- $k$  – liczba punktów pomiarowych przypisanych do krawędzi reprezentującej dany odcinek drogi w wybranym oknie czasowym.

Jak można zauważyć, prędkość nominalna przypisana do danej krawędzi została uwzględniona z wagą równą 2. Ma to na celu zapobiec sytuacjom, w których pojedyncze pomiary prędkości, mocno odstające od przypisanej wartości nominalnej, mają istotny wpływ na określenie przepustowości danego odcinka sieci drogowej.

Aktualizacja wag krawędzi w opisany powyżej sposób, a następnie wyznaczanie optymalnej trasy przejazdu za pomocą algorytmu korzystającego ze zmodyfikowanych wartości sprawiała, że podczas poszukiwania najszybszej trasy uwzględniane były informacje o aktualnym natężeniu ruchu drogowego.



Rys. 3: Liczba postojów o danej długości w analizowanej próbie danych

### 2.2.5 Wykrywanie postojów

Dane poddawane analizie opisywały przejazdy, które mogły zawierać postoje. Sytuacja, w której kilka odrębnych przejazdów opisanych jest za pomocą tego samego identyfikatora, była niepożądana z punktu widzenia prowadzonej analizy – mogłaby zaburzać otrzymane rezultaty.

Dysponując danymi w formie kolejnych punktów pomiarowych opisujących aktualne położenie, postój został zdefiniowany jako grupa kolejnych punktów (co najmniej dwóch), które znajdują się wewnątrz kwadratu o boku długości  $\sim 10$  metrów i środka wyznaczonym przez pierwszy punkt tworzący grupę. Do wyszukiwania tak zdefiniowanych grup punktów posłużył skrypt napisany w języku R. Znalezione postoje przeanalizowano pod kątem ich długości, by w ten sposób wyodrębnić zatrzymania spowodowane charakterystyką ruchu drogowego od zamierzonych przerw w jeździe. Te drugie, w tym przypadku, powinny być traktowane jako zdarzenia rozdzielające przejazd opisany tym samym identyfikatorem na niezależne przejazdy bez postojów.

Wykres (Rys. 3) pokazuje jak w analizowanej próbie przejazdów rozkłada się liczba znalezionych postojów w zależności od długości ich trwania. Długość trwania została podzielona na przedziały ze skokiem co 10 sekund.

Można zaobserwować, że dla początkowych grup liczba postojów maleje wraz ze wzrostem długości postoju. Tendencja ta jest jednak złamana począwszy od postojów o długości w przedziale 60 – 70 sekund. Kolejno obserwujemy wyraźny wzrosty liczby postojów o długości w przedziale 100 – 110 sekund i powolny, ale systematyczny spadek w kolejnych przedziałach.

Taka obserwacja pomogła w ustaleniu maksymalnej długości postoju, który będzie traktowany jako zatrzymanie wynikające ze specyfiki ruchu drogowego. W dalszych analizach przyjęto, że taką wartością jest zatrzymanie nie dłuższe niż 60 sekund. Zatrzymanie dłuższe niż 1 minuta będzie traktowane jako planowany postój i jeśli takowe zostanie wykryte dla przejazdu, który w analizowanych danych jest opisany jednym identyfikatorem, to zostanie on podzielony na niezależne przejazdy, które będą analizowane jako dwa zupełnie odrębne byty. Dodatkowo, określone w ten sposób założenie jest zgodne z definicjami zatrzymania i postoju, zawartymi w obowiązujących przepisach ruchu drogowego [24].



# Rozdział 3

## Analiza danych

Rozdział ten szczegółowo opisuje zagadnienia, które są głównym tematem niniejszej pracy magisterskiej. Przedstawiono w nim przeprowadzone analizy dostępnych danych ruchowych pod kątem znalezienia pewnych cech i wzorców, dzięki którym wyznaczane trasy będą możliwie najbliższe rzeczywistym decyzjom kierowców w ruchu miejskim.

### 3.1 Czy kierowcy jeżdżą trasami optymalnymi?

Pierwszym zagadnieniem związanym z analizą danych opisujących ruch drogowy, które zostało zbadane, było sprawdzenie, czy kierowcy wybierają trasy optymalne. Aby znaleźć odpowiedź na tak postawione pytanie, rozpatrzono w jakim stopniu rzeczywisty przejazd pokrywa się, z wyznaczoną na tym odcinku, drogą optymalną.

Analiza została rozpoczęta od przejazdów zarejestrowanych w Poznaniu i okolicach. Zbadana została próbka 525 przejazdów. Dla tej grupy kierowców wyliczono, że średnie pokrycie wybranej drogi z optymalną wynosiło 45%. Taki wynik był zastanawiająco niski i skłonił do tego, aby bardziej wnikliwie przyjrzeć się danym opisującym analizowane przejazdy. Po dokładniejszym zbadaniu wybranych przypadków okazało się, że kolejne punkty pomiarowe opisujące dany przejazd, oznaczone takim samym identyfikatorem przejazdu, de facto mogą być grupą kilku niezależnych przejazdów.

#### 3.1.1 Analiza przejazdów bez postojów

Sposób wykrywania postojów oraz podziału przejazdów na przejazdy bez parkowania, szczegółowo opisany w części 2.2.5, został zaaplikowany na dane poddawane analizie, po czym na wybranej wcześniej grupie zostało jeszcze

	Średnia	Wariancja
Przejazdy z postojami	0,45	0,13
Przejazdy bez postojów	0,62	0,11

Tab. 1: Pokrycie z trasą optymalną

raz sprawdzone, w jakim stopniu kierowcy poruszają się po trasach optymalnych. Teraz analizowanych było 538 przejazdów (pod uwagę brane były tylko te składające się z więcej niż 10 punktów pomiarowych) i zgodnie z przewidywaniami zaobserwowano dość znaczny wzrost średniego pokrycia trasy przejazdu z wyznaczoną trasą optymalną. Analizując przejazdy bez postojów, metryka ta przyjmowała wartość 62% (Tab. 1).

Takie rezultaty potwierdziły, że zaproponowany mechanizm wykrywania postojów i czyszczenia danych jest w powyższym przypadku właściwy, dlatego przyjęto, że wszystkie kolejne analizy będą przeprowadzane na uprzednio przygotowanych w ten sposób danych, czyli, że badane przejazdy nie będą zawierały postojów dłuższych niż jedna minuta.

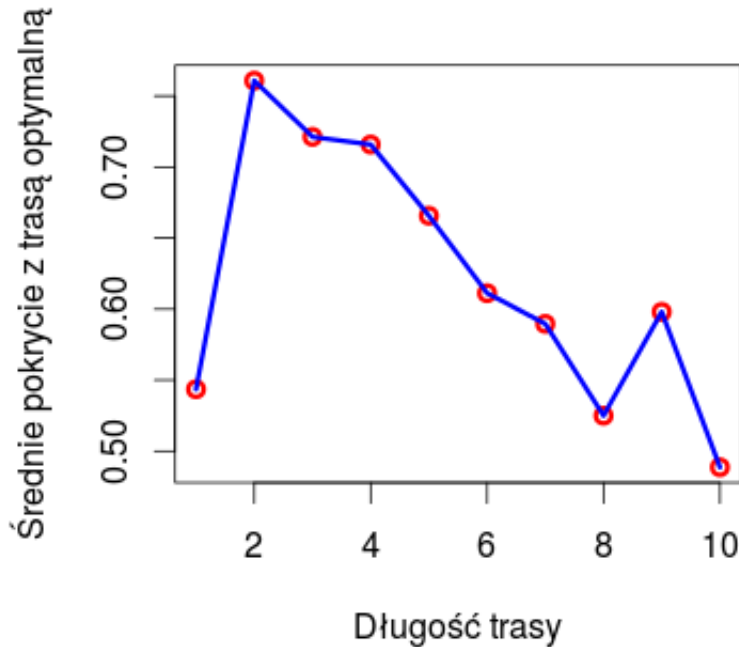
### 3.1.2 Długość trasy a pokrycie z trasą optymalną

Po obliczeniu średniej wartości określającej w jakim stopniu kierowcy decydują się wybierać trasy optymalne, ciekawym wydawało się sprawdzenie jak ta wartość zmienia się w zależności od pewnych cech charakteryzujących przejazdy. Pierwszym aspektem jaki został poddany analizie była długość trasy. Badana próbka 538 przejazdów bez postojów z Poznania i okolic została podzielona na 10 grup, każda o takiej samej liczbie elementów, rosnąco według długości. Zależność pomiędzy średnim pokryciem z trasą optymalną, a długością przejazdu zaprezentowana została na wykresie (Rys. 4). Na jego podstawie ciężko o jakieś daleko idące wnioski. Można jednak zauważyć, że dla tras dłuższych kierowcy w mniejszym stopniu jeżdżą trasami optymalnymi. Warto jednak zwrócić uwagę, że podobny wniosek możemy sformułować dla grupy 10% najkrótszych analizowanych przejazdów. Takie rezultaty mogą również wynikać ze stosunkowo niewielkich rozmiarów próby danych. W dalszej części pracy analizie poddawane będą zdecydowanie większe ilości danych.

### 3.1.3 Trasa optymalna – najkrótsza czy najszybsza?

W poprzednich eksperymentach trasa optymalna była definiowana jako najszybsza pod względem czasu przejazdu, gdyż takie podejście wydaje się





Rys. 4: Średnie pokrycie z trasą optymalną w zależności od długości trasy

najbardziej naturalne w kontekście ruchu drogowego. Interesującym wydawało się jednak również, jak zmieniłyby się wyniki, gdyby trasę optymalną rozpatrywać pod kątem długości, czyli szukać najkrótszej drogi. Przeanalizowano tę samą próbkę 538 przejazdów. Średnie pokrycie trasy przejazdu z trasą optymalną, rozumianą jako najkrótsza w sensie odległości, zmalało do 55%. Takie wyniki mogą być traktowane jako potwierdzenie hipotezy, że jednak w kontekście wyboru trasy przejazdu kierowcom częściej zależy na optymalizacji czasu przejazdu, a nie długości trasy. W dalszej części pracy skupiono się już tylko na rozpatrywaniu trasy optymalnej pod względem minimalizacji czasu przejazdu.

## 3.2 Analiza uwzględniająca natężenie ruchu

Kolejnym analizowanym zagadnieniem była kwestia wyznaczania tras optymalnych przy uwzględnianiu informacji o aktualnym natężeniu ruchu drogowego.

Eksperymenty rozpoczęto od sprawdzenia w jakim stopniu trasy przejazdów pokrywają się z trasami optymalnymi, wyznaczonymi z uwzględnieniem natężenia ruchu drogowego. Jako testową próbkę znów wykorzystano, tę samą co w poprzednich testach, grupę 538 przejazdów bez postojów z Poznania i okolic.

Wyznaczenie tras optymalnych w nowych warunkach sprawiło, że średnie pokrycie tras rzeczywistych z trasami optymalnymi wzrosło o dwa punkty procentowe i wyniosło w tym przypadku 64%. Jest to niewielka zmiana, ale zgodna z zakładaną hipotezą.

Warto jednak zaznaczyć, że na tym etapie badań sposób modelowania aktualnego natężenia ruchu drogowego był zdecydowanie mniej precyzyjny niż opisano to w części 2.2.4. W tym przypadku wartości przypisane do grafu reprezentującego sieć dróg zostały po prostu zmodyfikowane zgodnie ze wzorem 2.2, ale bez podziału przejazdów na okna czasowe, tylko jednorazowo, na podstawie danych z całej badanej próbki. Wynikało to między innymi ze zbyt małych rozmiarów próby, a także z faktu, że było to pierwsze podejście do modelowania aktualnego natężenia ruchu drogowego na potrzeby tej pracy. Ten aspekt pracy ewoluował i podczas kolejnych analiz był udoskonalany.

### **3.3 Dlaczego miałbym rezygnować ze swojej ulubionej trasy?**

Jedną z cech charakterystycznych dla ruchu miejskiego jest to, że spora grupa użytkowników dróg regularnie porusza się pomiędzy stałymi punktami, na przykład codzienna podróż z domu do pracy i z powrotem. Biorąc powyższe pod uwagę, można przypuszczać, że spora grupa kierowców w mieście zazwyczaj porusza się po swoich ulubionych, znanych im trasach. Dopiero informacja o poważnych utrudnieniach na wybranej przez nich drodze i związane z tym spore opóźnienia sprawiają, że są skłonni zmienić swoje przyzwyczajenia i zdecydować się na jazdę inną trasą, która w danych warunkach będzie zdecydowanie szybsza.

W celu zbadania postawionej hipotezy, kierowców podzielono na dwie grupy: tych, którzy uwzględniają informacje o aktualnym natężeniu ruchu drogowego oraz pozostałych. W tym celu dla każdego przejazdu zostały wyznaczone dwa warianty trasy optymalnej: pierwszy standardowy  $P_{B_i}$ , a drugi uwzględniający informację o aktualnym natężeniu ruchu drogowego  $P_{T_i}$ . Następnie zostało obliczone pokrycie  $C$  rzeczywistej trasy przejazdu z każdym z wariantów trasy optymalnej. Jeśli  $C$  dla wariantu trasy  $P_{T_i}$  było większe niż dla  $P_{B_i}$ , to taki przejazd był kwalifikowany do grupy przejazdów, dla których

Próbka	Mediana w gr. B	Mediana w gr. T
Poznań	4,63	22,69
Warszawa	7,49	23,78

Tab. 2: Mediana oszczędności czasu (w sekundach) z poruszania się trasą  $P_{T_i}$  w porównaniu do poruszania się trasą  $P_{B_i}$   
*gr. B* – kierowcy, którzy nie uwzględniają informacji o natężeniu ruchu drogowego  
*gr. T* – kierowcy, którzy uwzględniają informacje o natężeniu ruchu drogowego

Próbka	Średnia w gr. B	Średnia w gr. T
Poznań	17,15	42,74
Warszawa	13,73	37,55

Tab. 3: Średnia arytmetyczna oszczędności czasu (w sekundach) z poruszania się trasą  $P_{T_i}$  w porównaniu do poruszania się trasą  $P_{B_i}$   
*gr. B* – kierowcy, którzy nie uwzględniają informacji o natężeniu ruchu drogowego  
*gr. T* – kierowcy, którzy uwzględniają informacje o natężeniu ruchu drogowego

kierowcy uwzględniają wiedzę o aktualnym natężeniu ruchu.

W stawianej hipotezie występuje pojęcie zysku z wykorzystywania informacji o aktualnym natężeniu ruchu drogowego. Na potrzeby tej analizy zostało ono zdefiniowane jako różnica w teoretycznym czasie przejazdu, obliczonym przy użyciu mechanizmów biblioteki GraphHopper, pomiędzy wariantem trasy  $P_{T_i}$  a  $P_{B_i}$ .

Powyższe założenia miały zostać zweryfikowane poprzez analizę danych opisujących ruch drogowy w Poznaniu i w Warszawie w godzinach szczytu. Jako godziny o wzmożonym natężeniu ruchu zostały wybrane dwa okresy w ciągu dnia: 8:00 – 8:30 oraz 16:45 – 17:30, podzielone na piętnastominutowe okna czasowe. Spośród wszystkich dostępnych danych, wybrane zostały tylko te przejazdy, dla których wyznaczone warianty tras optymalnych były różne. Szczegółowe wyniki tej analizy przedstawione zostały w Dodatku A, natomiast podsumowanie przedstawiono tabelach (Tab. 2, Tab. 3, Tab. 4).

Na podstawie otrzymanych wyników można wnioskować, że zakładana hipoteza znajduje odzwierciedlenie w rzeczywistości. Dla analizowanej grupy przejazdów potwierdzają to wszystkie z obliczonych metryk. Dla grupy kie-

Próbka	Śr. zysk w p. proc. w gr. B	Śr. zysk w p. proc. w gr. T
Poznań	3,02	8,35
Warszawa	4,41	9,71

Tab. 4: Stosunek zysku czasu z poruszania się trasą  $P_{T_i}$  w porównaniu do poruszania się trasą  $P_{B_i}$  do czasu długości trwania przejazdu wyrażony w punktach procentowych

*gr. B* – kierowcy, którzy nie uwzględniają informacji o natężeniu ruchu drogowego

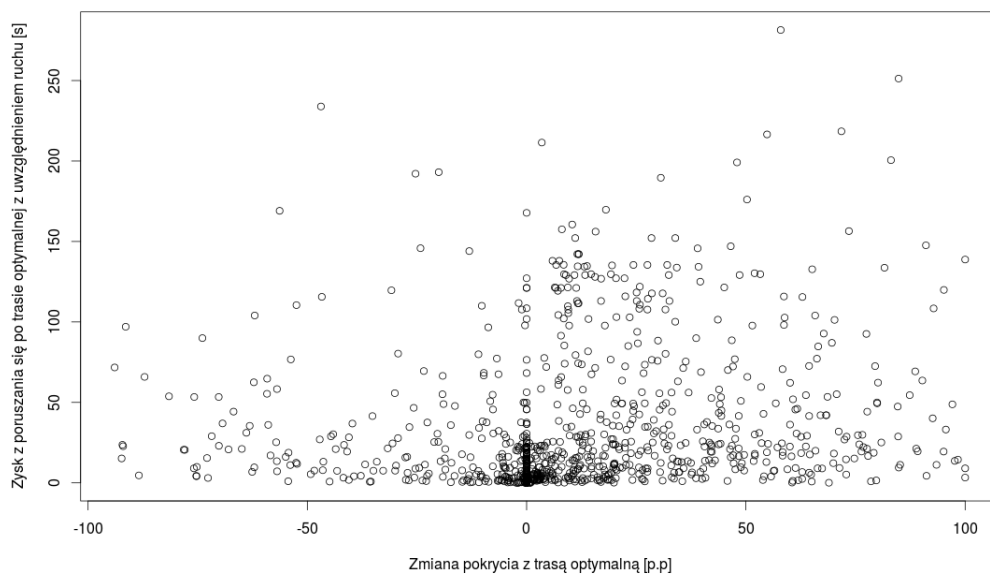
*gr. T* – kierowcy, którzy uwzględniają informacje o natężeniu ruchu drogowego

rowców uwzględniających informacje o aktualnym natężeniu ruchu drogowego zdecydowanie wyższa jest zarówno średnia, jak i mediana zysku, wyrażonego jako oszczędność czasu wynikająca z wyboru trasy wyznaczonej z uwzględnieniem aktualnej sytuacji na drogach. Również średni zysk wyrażony jako stosunek oszczędzonego czasu do całkowitego czasu przejazdu jest wyższy w grupie kierowców biorących pod uwagę dostępne informacje o natężeniu ruchu. Powyższe pokazuje, iż można zakładać, że im więcej czasu pozwoli zaoszczędzić kierowcy uwzględnienie dynamicznie zmieniającej się sytuacji na drogach podczas wyboru trasy, tym chętniej będzie on z takich informacji korzystać.

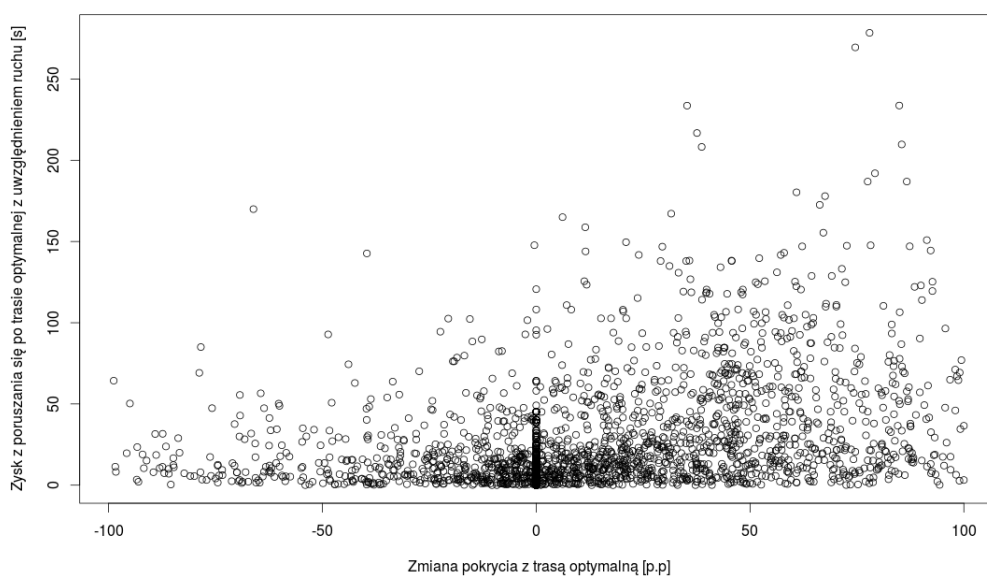
Zebrane podczas tej analizy wyniki, w formie graficznej zostały zaprezentowane na wykresach (Rys. 5 i Rys. 6). Także tutaj możemy zaobserwować, że punkty na prawo od początku układu współrzędnych na osi odciętych, czyli te reprezentujące grupę kierowców uwzględniających informacje o natężeniu ruchu, mają generalnie większe wartości na osi rzędnych, która reprezentuje oszczędność czasu. Szczególnie dobrze widać to na wykresie prezentującym dane z Warszawy (Rys. 6).

### 3.4 Analiza ruchu w godzinach szczytu

Po potwierdzeniu hipotezy, że im większa oszczędność czasu wynikająca z uwzględniania informacji o bieżącym ruchu drogowym, tym chętniej kierowcy z takiej informacji korzystają, warto było przyjrzeć się otrzymanym rezultatom również pod innym kątem.



Rys. 5: Zmiana pokrycia z trasą optymalną w stosunku do zysku czasu wynikającego z tego faktu dla danych z Poznania



Rys. 6: Zmiana pokrycia z trasą optymalną w stosunku do zysku czasu wynikającego z tego faktu dla danych z Warszawy

### **3.4.1 Szczegóły analizowanego zbioru przejazdów**

Szczegóły ilościowe opisujące analizowane dane przedstawiają się następująco:

- Poznań
  - 4319 przejazdów w godzinach 8:00-8:30 i 16:45-17:30,
  - dla 1081 (427 przejazdów rano i 654 przejazdów popołudniu) z 4319 przejazdów, wyznaczenie trasy optymalnej z uwzględnieniem ruchu drogowego daje nową trasę optymalną (w stosunku do trasy wyznaczonej bez uwzględnienia ruchu drogowego),
  - 576 z 1081 (53 procent) kierowców uwzględnia informacje o ruchu drogowym,
  
- Warszawa
  - 8516 przejazdów w godzinach 8:00-8:30 i 16:45-17:30,
  - dla 2462 przejazdów (1034 przejazdów rano i 1428 przejazdów popołudniu) z 8516 przejazdów, wyznaczenie trasy optymalnej z uwzględnieniem ruchu drogowego daje nową trasę optymalną (w stosunku do trasy wyznaczonej bez uwzględnienia ruchu drogowego),
  - 1314 z 2462 (53 procent) kierowców uwzględnia informacje o ruchu drogowym.

### **3.4.2 Pokrycie z trasą optymalną**

Pierwszym aspektem było sprawdzenie czy kierowcy jeżdżą zgodnie z wyznaczonymi trasami optymalnymi, czyli powrót do zagadnienia, od którego rozpoczęła się analiza danych, tylko, że tym razem na dużo większej próbie danych i z bardziej dopracowanym sposobem modelowania aktualnego natężenia ruchu drogowego. Szczegółowe wyniki przedstawiono w tabeli (Tab. 5). Na ich podstawie można sformułować wniosek, że nawigacja z uwzględnieniem aktualnego ruchu drogowego jest bliższa rzeczywistym trasom wybieranym przez kierowców. Przyjmując jako miarę część wspólną trasy przejazdu z trasą optymalną (pokrycie  $C$ ), obserwujemy wzrost pokrycia w przypadku, gdy uwzględniamy podczas wyznaczania trasy optymalnej ruch drogowy (wzrost średniej i mediany, spadek wariancji).

	Średnia	Mediana	Wariancja
Kierowcy nieuwzględniający informacji o bieżącym ruchu (Poznań)	0,78	0,94	0,09
Kierowcy uwzględniający informacje o bieżącym ruchu (Poznań)	0,8	0,96	0,08
Kierowcy nieuwzględniający informacji o bieżącym ruchu (Warszawa)	0,77	0,95	0,09
Kierowcy uwzględniający informacje o bieżącym ruchu (Warszawa)	0,81	0,96	0,07

Tab. 5: Pokrycie z trasą optymalną

Próbka	Długość trasy > 10 km	Długość trasy ≤ 10 km
Poznań	54%	53%
Warszawa	67%	49%

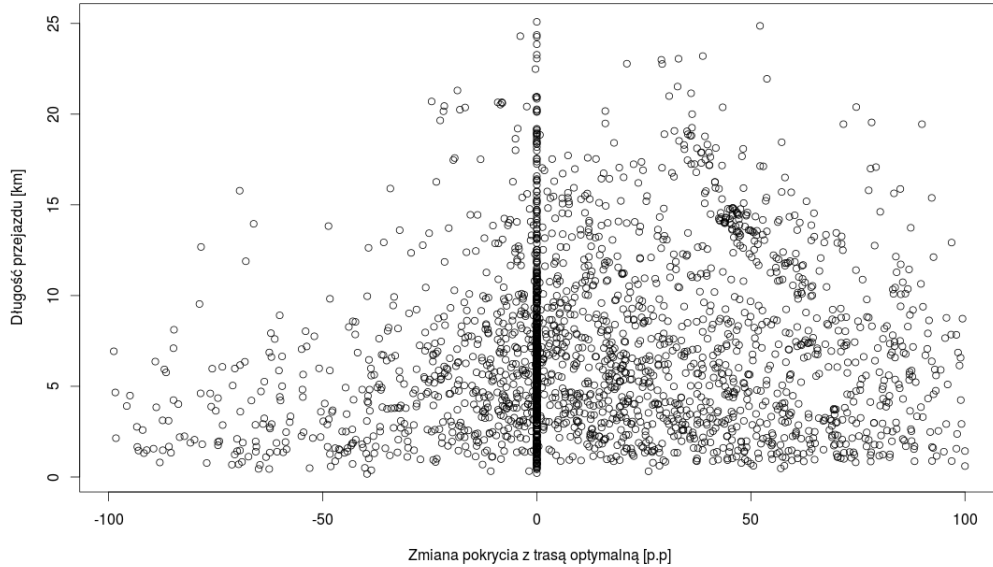
Tab. 6: Odsetek kierowców uwzględniających informacje o bieżącym natężeniu ruchu w zależności od długości trasy

### 3.4.3 Długość trasy

Kolejną kwestią było zbadanie zależności pomiędzy długością trasy przejazdu, a tym jak chętnie kierowcy korzystają z informacji o bieżącym natężeniu ruchu. W tym celu przejazdy zostały podzielone na dwie grupy: dłuższe niż 10 kilometrów i pozostałe. Tabela (Tab. 6) przedstawia jaka część kierowców w tych dwóch grupach korzystała z informacji o aktualnym ruchu na drogach.

W otrzymanych wynikach ciekawy jest fakt, że dla analizowanej próbki danych z Poznania, podział przejazdów według ich długości nie ma wpływu na to jaka część kierowców w danej grupie uwzględnia informacje o bieżącym ruchu drogowym. Zupełnie inaczej przedstawiają się natomiast wyniki dla danych z Warszawy. W tym przypadku różnica jest znaczna i obserwujemy, że kierowcy, poruszając się na długich trasach, zdecydowanie chętniej korzystają z informacji o bieżącym natężeniu ruchu. Pokazuje to także wyraźnie wykres (Rys. 7), gdzie zdecydowana większość punktów odpowiadających przejazdom dłuższym niż 10 km leży na prawo od osi  $y$ .

Takie wnioski wpisują się w tezę, że im potencjalnie większa korzyść z poruszania się trasą optymalną wyznaczoną z uwzględnieniem aktualnej sytuacji na drogach, tym kierowcy chętniej z niej korzystają. Generalnie im trasa jest dłuższa, tym zazwyczaj taki zysk może być większy i wyniki uzyskane na podstawie analizy danych z Warszawy są potwierdzeniem tego założenia.



Rys. 7: Zmiana pokrycia z trasą optymalną w stosunku do długości trasy przejazdu dla danych z Warszawy

Próbka	8:00 – 8:30	16:45 – 17:30
Poznań	55%	52%
Warszawa	56%	51%

Tab. 7: Odsetek kierowców uwzględniających informacje o bieżącym natężeniu ruchu w zależności od pory dnia

Warto dodać, że różnica w wynikach pomiędzy danymi z Warszawy i Poznania może wynikać z faktu, iż dane z Warszawy pochodzą z lutego 2015 roku, natomiast dane z Poznania datowane są na luty 2014 roku. Ten rok różnicy może mieć istotny wpływ na upowszechnienie się systemów nawigacji uwzględniających informacje o natężeniu ruchu w czasie rzeczywistym.

### 3.4.4 Pora dnia

Analizowane dane podzielone zostały także pod względem pory dnia, w jakiej przejazd miał miejsce. Pod uwagę bierzemy dwa szczyty komunikacyjne: poranny (8:00 – 8:30) i popołudniowy 16:45 – 17:30. Tabela (Tab. 7) przedstawia jaką część kierowców w tych grupach uwzględnia informacje o aktualnej sytuacji na drogach podczas wyboru trasy przejazdu.



Zysk z poruszania się nową trasą optymalną (w sekundach)	Średnia zmiana pokrycia z trasą optymalną	Liczba przejazdów
(0; 25]	0,07	720
(25; 50]	0,19	148
(50; 100]	0,16	99
(100; 150]	0,22	92
> 150	0,24	22

Tab. 8: Zmiana pokrycia z trasą optymalną w zależności od skrócenia czasu przejazdu dzięki poruszaniu się trasą o  $P_{T_i}$  dla danych z Poznania

Otrzymane rezultaty pokazują, że informacje o aktualnym natężeniu ruchu są chętniej brane pod uwagę przy planowaniu trasy przejazdu podczas porannego szczytu komunikacyjnego. Fakt ten można tłumaczyć na przykład tym, że rano kierowcom częściej zależy na punktualności (dojazd na czas do pracy, szkoły, itp.) niż popołudniu, gdy większość uczestników ruchu miejskiego wraca spokojnie do domu. W tym przypadku, różnice w otrzymanych wynikach są jednak na tyle nieznaczące, że ciężko wysuwać na ich podstawie jakieś daleko idące wnioski.

### 3.4.5 Skrócenie czasu przejazdu

Wcześniejsze analizy potwierdziły hipotezę, iż im większe skrócenie czasu przejazdu dzięki uwzględnieniu informacji o aktualnym ruchu na drogach, tym chętniej kierowcy biorą te informacje pod uwagę przy wyborze trasy. Kolejnym elementem weryfikowania powyższej hipotezy jest podział przejazdów według kryterium oszczędności czasu z poruszania się trasą  $P_{T_i}$  w stosunku do poruszania się trasą  $P_{B_i}$ .

Wszystkie analizowane przejazdy, dla których wyznaczone trasy optymalne w tych dwóch wariantach były różne, zostały przypisane do jednej z pięciu grup na podstawie tego jak duże było skrócenie czasu przejazdu po uwzględnieniu informacji o natężeniu ruchu. Dla każdej z tych grup wyliczono jaka jest średnia zmiana pokrycia  $C$  z trasą  $P_{T_i}$  w porównaniu do pokrycia z wariantem  $P_{B_i}$ . Szczegółowe wyniki na podstawie analizy danych z Poznania i Warszawy przedstawiono odpowiednio tabelach: Tab. 8 i Tab. 9.

W otrzymanych rezultatach wyraźnie widać prawidłowość, że wraz ze skróceniem czasu przejazdu, wynikającym z uwzględnienia informacji o natężeniu ruchu, wzrasta różnica w pokryciu  $C$  pomiędzy wariantami  $P_{T_i}$  i  $P_{B_i}$  (na

Zysk z poruszania się nową trasą optymalną (w sekundach)	Średnia zmiana pokrycia z trasą optymalną	Liczba przejazdów
(0; 25]	0,06	1645
(25; 50]	0,22	390
(50; 100]	0,37	320
(100; 150]	0,45	88
> 150	0,53	19

Tab. 9: Zmiana pokrycia z trasą optymalną w zależności od skrócenia czasu przejazdu dzięki poruszaniu się trasą  $P_{T_i}$  dla danych z Warszawy

korzystać tego pierwszego). Szczególnie wyraźnie widać to dla danych z Warszawy, gdzie dla grupy przejazdów z nowo wyznaczoną trasą optymalną szybszą o nie więcej niż 25 sekund, średnie pokrycie z trasą optymalną wzrasta tylko o 6 punktów procentowych. Natomiast dla dwóch grup, gdzie skrócenie czasu przejazdu najszybszą trasą jest z przedziału (100; 150] sekund czy powyżej 150 sekund, wzrost średniego pokrycia oscyluje w okolicach 50 punktów procentowych.

### 3.4.6 Pożar mostu Łazienkowskiego i jego wpływ na ruch drogowy w Warszawie

14 lutego 2015 roku doszło do pożaru części mostu Łazienkowskiego w Warszawie, czego skutkiem było jego wyłączenie z użycia w okresie od 14 lutego 2015 do 28 października 2015 roku [25]. Dostępność danych opisujących ruch drogowy w Warszawie w dniach 12 i 19 lutego 2015, umożliwiła analizę jaki wpływ na wybór trasy przez kierowców, pod kątem zgodności z wyznaczoną trasą optymalną, ma zamknięcie jednej z kluczowych dróg w mieście. W tabeli (Tab. 10) zaprezentowano wyniki analizy z podziałem przejazdów z Warszawy według dnia (przed i po zamknięciu mostu Łazienkowskiego).

W tym przypadku otrzymane rezultaty raczej nie pozwalają na wyciąganie jakichś wiążących wniosków. Może to być spowodowane tym, że analizowana próbka przejazdów była zbyt ogólna i zamknięcie mostu Łazienkowskiego wpłynęło tylko na niewielką część z tych przejazdów. Dodatkowo wyniki mógł zaburzać fakt, że podczas wyznaczania tras optymalnych dla przejazdów z dnia 19 lutego 2015 r. nie zostało uwzględnione wyłączenie z użycia mostu Łazienkowskiego. Oba te problemy zostały zaadresowane podczas dalszej analizy.

	12.02.2015 r.	19.02.2015 r.
Ogółem	53%	54%
Na trasach dłuższych niż 10 km	67%	67%
Na trasach nie dłuższych niż 10 km	49%	50%
Rano (8:00 – 8:30)	58%	55%
Popołudniu (16:45 – 17:30)	49%	54%

Tab. 10: Odsetek kierowców uwzględniających informacje o ruchu drogowym

### 3.5 Analiza ruchu na warszawskich mostach

W podrozdziale 3.4.6 porównywano charakterystyki przejazdów odbywających się w godzinach szczytu w Warszawie w podziale na okres sprzed i po zamknięciu mostu Łazienkowskiego. Analiza ta nie pozwoliła jednak na sformułowanie istotnych wniosków. Potencjalnie przyczyną takiego stanu rzeczy mogła być zbyt mała wielkość próbki, która została poddana badaniu. Wtedy pod uwagę brane były tylko przejazdy, które miały miejsce w godzinach porannego albo popołudniowego szczytu komunikacyjnego. W dalszej opisanych eksperymentach analizę skupiono na tych przejazdach z całej doby, których trasa przebiega przez którąś z przepraw na Wiśle w Warszawie.

Tak postawione zadanie wymagało rozbudowania aplikacji służącej do wyznaczania tras optymalnych o funkcjonalność wykrywania przejazdów przez mosty. Kolejną modyfikacją jaka została wprowadzona w aplikacji na potrzeby tego eksperymentu, była możliwość wyłączania z użycia zadanych odcinków ulic podczas wyznaczania trasy optymalnej. Dzięki tej funkcji możliwe było odwzorowanie sytuacji, gdzie dla przejazdów, które miały miejsce w dniu 19 lutego 2015 r., most Łazienkowski był zamknięty – zostało to uwzględnione przez algorytm wyznaczający optymalną trasę przejazdu.

#### 3.5.1 Analiza różnych wariantów tras optymalnych

W badanej próbie danych, opisujących ruch drogowy w Warszawie w dniach 12 i 19 lutego 2015 roku, znalezione zostały 10374 przejazdy, dla których wyznaczenie trasy optymalnej z uwzględnieniem ( $P_{T_i}$ ) i bez uwzględnienia aktualnego natężenia ruchu drogowego ( $P_{B_i}$ ) skutkowało wyborem różnych tras przejazdu oraz kierowca przekraczał Wisłę poruszając się po którymś z warszawskich mostów. Spośród tak wyselekcjonowanych danych wyodrębnione zostały tylko te przypadki, dla których każdy z dwóch wariantów trasy optymalnej zalecał przekroczenie Wisły po innym moście. W analizowanej próbce taka sytuacja miała miejsce 1555 razy.

Mając tak przygotowane dane, możliwa była segmentacja kierowców według kolejnego kryterium – po jakim moście odbywa się przejazd. Przy tak postawionym zagadnieniu, każda z tras przejazdów mogła prowadzić:

- albo po moście, po którym prowadzi trasa  $P_{B_i}$  („Blind”),
- albo po moście, po którym prowadzi trasa  $P_{T_i}$  („Traffic”),
- albo po dowolnym innym moście w Warszawie („Other”).

Przyjmując takie kryterium możliwy był podział, podobnie jak w poprzednich eksperymentach, na kierowców, którzy wybierając trasę uwzględniają informacje o aktualnej sytuacji na drogach i pozostałych. Różnica polegała na tym, że w poprzednich analizach o zaklasyfikowaniu kierowcy do grupy uwzględniającej informacje o bieżącej sytuacji drogowej decydował już nawet minimalny wzrost współczynnika  $C$ , w przypadku trasy optymalnej uwzględniającej aktualne warunki drogowe. Tak przyjęta miara mogła powodować zbyt pochopne klasyfikowanie kierowców jako uwzględniających informacje o ruchu drogowym (tzw. *false positives*).

W warunkach warszawskich, gdzie przepraw drogowych na Wiśle nie ma zbyt wiele, wybór mostu w znacznym stopniu wpływa na całą trasę przejazdu. Dzięki temu podział ze względu na most, po którym odbywa się przejazd, pozwala na zaklasyfikowanie kierowców do grupy biorących pod uwagę natężenie ruchu drogowego podczas wyboru trasy w sposób pewniejszy (ograniczenie liczby *false positives*), niż miało to miejsce w poprzednich analizach.

W tabeli (Tab. 11) zaprezentowano zbiorcze wyniki pokazujące jak zmienia się stosunek liczby kierowców uwzględniających informacje o bieżącym natężeniu ruchu drogowego (przy zastosowaniu wcześniej opisanego kryterium podziału według mostów), w zależności od zysku czasu wynikającego z uwzględnienia tych informacji. Szczegółowe wyniki z podziałem na poszczególne dni są zaprezentowane w tabeli (Tab. 23) w Dodatku A.

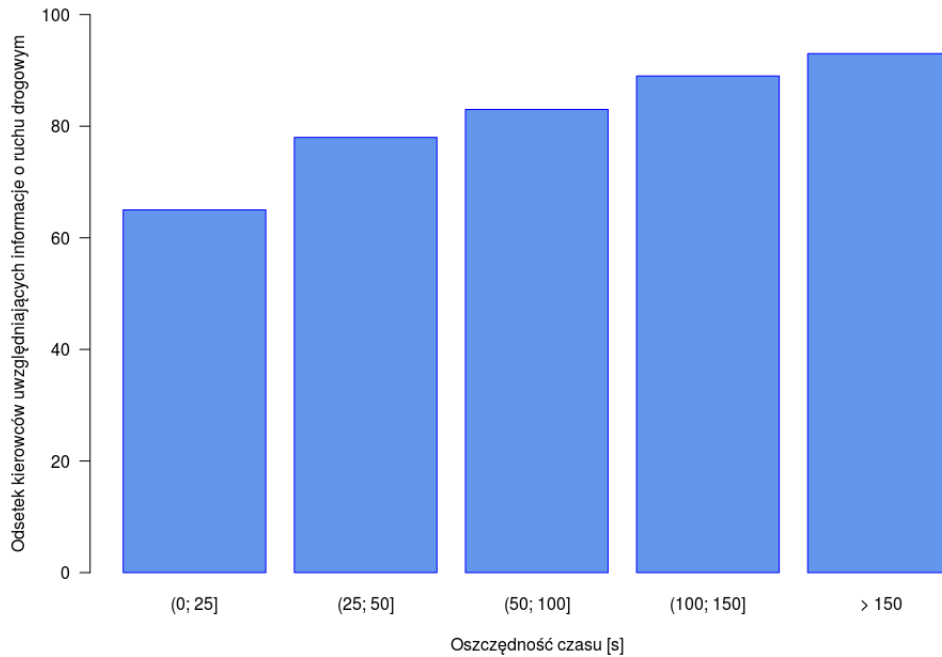
Otrzymane wyniki pokazują, iż kierowcy tym chętniej uwzględniają informacje o aktualnym natężeniu ruchu drogowego, im bardziej skrócony zostanie czas przejazdu do miejsca docelowego dzięki uwzględnieniu takich informacji. Taką tendencję widać bardzo wyraźnie, jeśli analizie poddamy tylko kierowców, których określamy jako korzystających z systemów nawigacji. W takim przypadku odrzucamy przejazdy zaklasyfikowane do grupy „Other”. Dla kolejnych przedziałów zysku czasu obserwujemy wtedy monotoniczny wzrost stosunku kierowców korzystających z nawigacji uwzględniającej natężenie ruchu w czasie rzeczywistym do wszystkich kierowców. Zależność ta została zaprezentowana na wykresie (Rys. 8).

Zysk z poruszania się nową trasą optymalną [w sekundach]	Odsetek kierowców uwzględniających informacje o ruchu drogowym $\frac{ T }{ B + T + O }$	Odsetek kierowców uwzględniających informacje o ruchu drogowym (po odrzuceniu kierowców niekorzystających z nawigacji) $\frac{ T }{ B + T }$
(0; 25]	54%	65%
(25; 50]	65%	78%
(50; 100]	64%	83%
(100; 150]	67%	89%
Powyżej 150	62%	95%

Tab. 11: Rezultaty pokazujące jaka część kierowców uwzględnia informacje o bieżącym natężeniu ruchu drogowego, w zależności od stopnia w jakim uwzględnienie takiej informacji skraca czas przejazdu, gdzie:

$|T|$  – liczba kierowców jadących przez most, po którym prowadzi trasa  $P_{T_i}$ ,  
 $|B|$  – liczba kierowców jadących przez most, po którym prowadzi trasa  $P_{B_i}$ ,  
 $|O|$  – liczba kierowców jadących przez most, po którym nie prowadzi ani trasa  $P_{T_i}$ , ani  $P_{B_i}$

Wnioski wyciągnięte na podstawie powyższej analizy wydają się być zgodne z intuicją. Trasa optymalna wyznaczona w wariancie „Blind” prowadzi po ulicach, po których nominalny czas przejazdu powinien być najkrótszy. Często są to główne, wielopasmowe arterie, o wysokiej przepustowości, z ograniczeniem prędkości podniesionym powyżej, zazwyczaj obowiązującego w terenie zabudowanym limitu 50 km/h. Jazda takim trasami powinna być szybka i wygodna, dlatego często jest naturalnym wyborem dla sporej grupy kierowców. To z kolei sprawia, że w godzinach szczytu są one mocno zatłoczone, czy wręcz zakorkowane i w takich warunkach nominalny czas przejazdu ma niewiele wspólnego z rzeczywistym. Dzięki systemom nawigacji uwzględniającym natężenie ruchu drogowego w czasie rzeczywistym, kierowcy są w stanie na bieżąco reagować na zmieniające się warunki na drogach. Informacja o tym, że poruszanie się drugorzędnymi drogami, które zazwyczaj są mniej komfortowe dla kierującego pojazdem, spowoduje nieznaczne skrócenie czasu



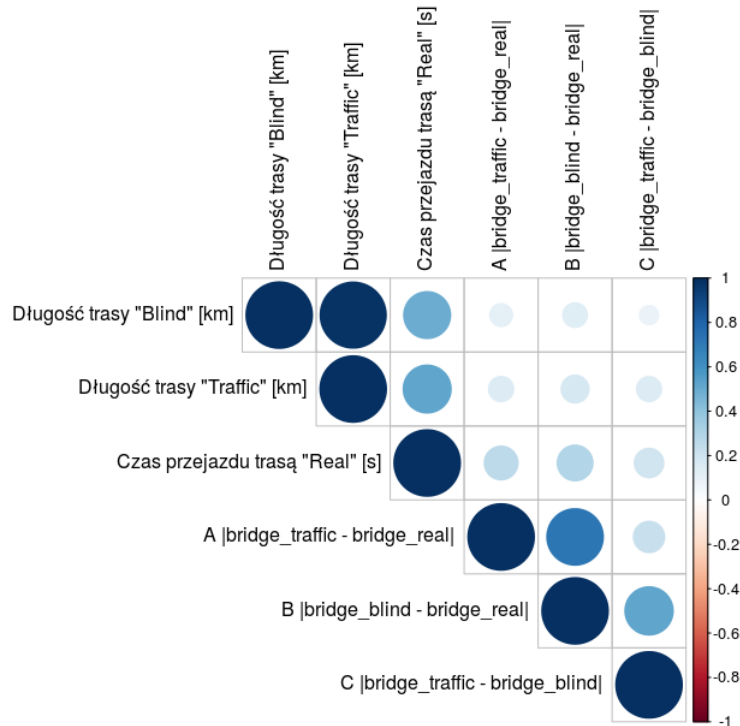
Rys. 8: Odsetek kierowców uwzględniających informacje o ruchu drogowym w kolejnych przedziałach skrócenia czasu przejazdu

przejazdu do miejsca docelowego, najprawdopodobniej nie zachęci większości kierowców do wyboru takiej trasy. Jednak im ta oszczędność czasu będzie większa, tym liczniejsza grupa planujących podróż będzie skłonna wybrać taką opcję. Otrzymane podczas tej analizy wyniki są zgodne z powyższą hipotezą.

### 3.5.2 Analiza korelacji

Przyjmując jako kryterium segmentacji most, po którym odbywał się przejazd („Blind”, „Traffic”, „Other”), zdefiniowano jeszcze jedną miarę charakteryzującą badane przejazdy – odległość od trasy optymalnej wyrażoną w mostach. Kolejnym warszawskim mostom, poczynając od mostu Siekierkowskiego, a kończąc na moście Północnym, przypisano identyfikatory od 1 do 8. Odległość wyrażoną w mostach zdefiniowano jako wartość bezwzględna z różnicy pomiędzy identyfikatorami mostów:

- $A|bridge\_traffic - bridge\_real|$  – odległość pomiędzy mostem na trasie  $P_{T_i}$  a mostem, po którym odbywał się rzeczywisty przejazd,
- $B|bridge\_blind - bridge\_real|$  – odległość pomiędzy mostem na trasie



Rys. 9: Wykres korelacji pomiędzy wybranymi cechami przejazdów

$P_{B_i}$  a mostem, po którym odbywał się rzeczywisty przejazd,

- $C|bridge\_traffic - bridge\_blind|$  – odległość pomiędzy mostem na trasie  $P_{T_i}$  a mostem na trasie  $P_{B_i}$ .

Tak przyjęta miara wykorzystana została w celu zbadania zachodzących korelacji. Wyznaczona została macierz korelacji (Tab. 24 w Dodatku A), której graficzna reprezentacja została przedstawiona na wykresie (Rys. 9).

Na podstawie otrzymanych rezultatów można sformułować wniosek, że wraz ze wzrostem długości trasy, zarówno liczonej jako długość wyrażona w pokonanej odległości, jak i jako czas, który był potrzebny na jej pokonanie, obserwujemy wzrost odległości wyrażonej w mostach pomiędzy rzeczywistą trasą przejazdu, a oboma wariantami trasy optymalnej. Fakt ten można tłumaczyć tym, że im dłuższa trasa, tym większy jest zakres możliwości, jeśli chodzi o wybór mostu. Formułując powyższy wniosek należy dodatkowo mieć na uwadze, że obserwowane korelacje są korelacjami słabymi (zazwyczaj przyjmuje się, że korelacje o wartości  $< 0,3$  opisują słaby związek).

## 3.6 Porównanie wybranych modeli klasyfikacyjnych

Ostatnią kwestią badaną w ramach niniejszej pracy była weryfikacja wybranych metod klasyfikacyjnych. Jako dane służące do przygotowania modeli wykorzystane zostały wybrane uprzednio przejazdy, których trasa wiodła, przez któryś z warszawskich mostów na Wiśle i dodatkowo trasy  $P_{B_i}$  oraz  $P_{T_i}$  przebiegały przez różne mosty. Zbiór tych przejazdów został losowo podzielony w stosunku 7:3, odpowiednio na zbiór treningowy i zbiór testowy. Do porównania zostały wybrane trzy powszechnie stosowane sposoby prognozowania zmiennej dyskretnej, a mianowicie:

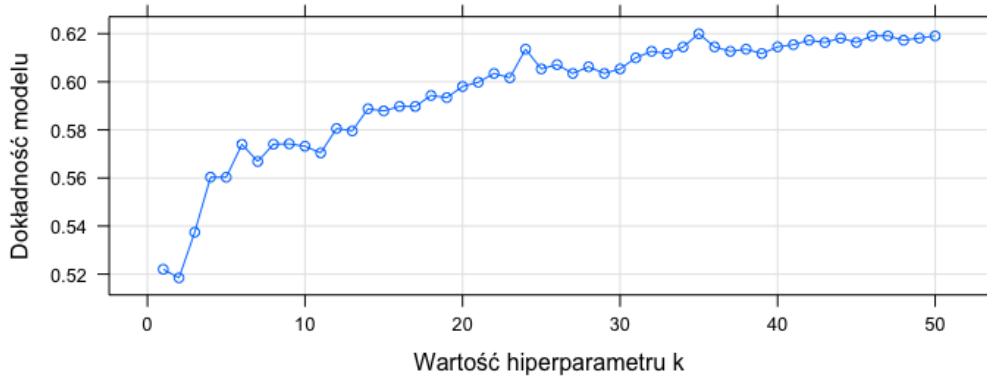
- metoda k najbliższych sąsiadów,
- metoda Naive Bayes,
- drzewa decyzyjne.

W stworzonych modelach zmienną objaśnianą jest rodzaj przejazdu i może przyjmować ona jedną z trzech wartości „Traffic”, „Blind” bądź „Other”, zgodnie z definicją opisaną w części 3.5.1. Zmiennymi objaśniającymi są natomiast:

- długość trasy  $P_{B_i}$ ,
- czas przejazdu trasą  $P_{B_i}$ ,
- długość trasy  $P_{T_i}$ ,
- czas przejazdu trasą  $P_{T_i}$ ,
- pora dnia, w której przejazd ma miejsce.

Wszystkie zmienne objaśniające są ciągłe, oprócz pory dnia, która może przyjmować jedną z pięciu wartości określających część dnia, w której odbywał się przejazd. Tak skonstruowany model może posłużyć do przewidywania trasy, jaką, w określonych warunkach drogowych, będzie poruszał się kierowca.





Rys. 10: Dokładność modelu k najbliższych sąsiadów w zależności od wartości hiperparametru k

### 3.6.1 Metoda k najbliższych sąsiadów

Metoda k najbliższych sąsiadów [26], jest to jeden z najprostszych algorytmów klasyfikacyjnych. Pomimo tego znajduje szerokie zastosowanie w wielu dziedzinach, takich jak na przykład: rozpoznawanie twarzy, przewidywanie upodobań czy wyszukiwanie odpowiednich sekwencji w kodzie DNA. Generalnie idea algorytmu polega na przydzieleniu obserwacji o nieznanym typie do jednego z typów, na podstawie jego najbliższych sąsiadów. Przy opracowywaniu modelu bazującego na metodzie k najbliższych sąsiadów, często ciężko jest o właściwy dobór wartości hiperparametru k, czyli właśnie liczby najbliższych sąsiadów uwzględnianych przez algorytm. Na wykresie (Rys. 10) zaprezentowano jak zmienia się dokładność (ang. *accuracy*) opracowanego modelu w zależności od wartości k.

Testowanie wartości k z zakresu (0; 50] pokazuje, że największą dokładność można osiągnąć dla  $k = 35$ . Dla tak ustawionego hiperparametru, model uruchomiony na zbiorze testowym działa z dokładnością wynoszącą 0,61. Jest to wartość nieznacznie niższa od osiągniętej na zbiorze treningowym, gdzie dokładność wynosiła 0,62. Z kolei w tabeli (Tab. 12) przedstawiono macierz błędów (ang. *confusion matrix*) otrzymaną w wyniku działania opracowanego modelu na zbiorze testowym.

Analizując otrzymaną macierz błędów oraz inne miary statystyczne powszechnie używane przy ocenie modelu, można dojść do wniosku, że opracowany model, pomimo akceptowalnego poziomu dokładności, jednak nie działa najlepiej. Wrażliwość (ang. *sensitivity*), czyli miara, która pokazuje jaka część obserwacji z danej klasy, została właściwie zaklasyfikowana, jest bardzo niska

		Obserwacje		
		Blind	Other	Traffic
Predykcje	Blind	0	0	0
	Other	4	4	8
	Traffic	79	91	280

Tab. 12: Macierz błędów modelu k najbliższych sąsiadów dla  $k = 35$

dla klas „Blind” oraz „Other”. Dla wyników na zbiorze testowym, wrażliwość dla poszczególnych klas ma następujące wartości:

- „Blind” – 0,
- „Traffic” – 0,97,
- „Other” – 0,04,

Wynik dla klasy „Traffic” jest bardzo dobry, ale wyniki dla klas „Blind” oraz „Other” są mocno niepokojące. Wrażliwość dla klasy „Blind” równa 0 oznacza, że żadna z obserwacji tej klasy nie została poprawnie zaklasyfikowana przez przygotowany model. W parze z wrażliwością często podaje się wartość swoistości (ang. *specificity*) modelu. Dla wyników otrzymanych na zbiorze testowym swoistość dla klasy „Traffic” wynosi 0,04, co oznacza, że bardzo wiele obserwacji innych klas jest niepoprawnie klasyfikowane przez model jako obserwacje klasy „Traffic” (dla klas „Blind i „Other” swoistość wynosi odpowiednio 1 i 0,97), co doskonale można zauważyć także w przedstawionej macierz błędów.

Powyższe wyniki skłoniły do dokładniejszej analizy modelu dla mniejszej wartości hiperparametru  $k$ . Wartość ta została ustalona na równą 5. Macierz błędów dla tego przypadku przedstawiono w tabeli (Tab. 13).

Dokładność modelu w tym wariantcie ( $k = 5$ ) na zbiorze testowym wynosi 0,55, ale z drugiej strony obserwowany jest wyraźny wzrost wrażliwości dla klas „Blind” i „Other” (w obu przypadkach 0,25). Ogólny spadek dokładności modelu pokazuje jednak, że generalnie w powyższym przypadku ciężko może być uzyskać model oparty na metodzie  $k$  najbliższych sąsiadów, który, na podstawie przyjętych parametrów, z dużym prawdopodobieństwem będzie poprawnie klasyfikował kierowców.

		Obserwacje		
		Blind	Other	Traffic
Predykcje	Blind	21	8	26
	Other	9	24	50
	Traffic	53	63	212

Tab. 13: Macierz błędów modelu k najbliższych sąsiadów dla  $k = 5$ 

		Obserwacje		
		Blind	Other	Traffic
Predykcje	Blind	2	0	0
	Other	5	20	21
	Traffic	76	75	267

Tab. 14: Macierz błędów modelu opartego o klasyfikator Naive Bayes

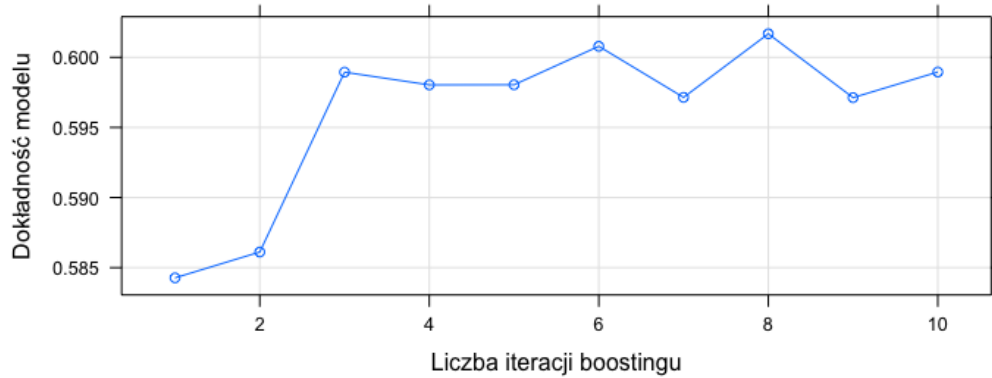
### 3.6.2 Metoda Naive Bayes

Wyniki uzyskane przy użyciu modelu klasyfikacyjnego opartego na metodzie k najbliższych sąsiadów można porównać z modelami opartymi na innych metodach, na przykład na metodach bayesowskich. Najprostszym klasyfikatorem bayesowskim jest metoda Naive Bayes. Najwyższą dokładność modelu jaką udało się osiągnąć, wykorzystując do doboru jak najlepszych parametrów modelu pakiet Caret środowiska R, wyniosła 0,62. Dokładniejsze rezultaty działania modelu na zbiorze testowym pokazuje macierz błędów zaprezentowana w tabeli (Tab. 14).

Powyższe rezultaty pokazują, że model oparty na klasyfikatorze Naive Bayes działa nieznacznie lepiej niż oparty na metodzie k najbliższych sąsiadów. Cechuje się on nieznacznie wyższą dokładnością, a ponadto wyższa jest również wrażliwość modelu dla klas „Blind” oraz „Other” (odpowiednio 0,02 i 0,21), chociaż dalej pozostaje ona daleka od pożądanych wartości.

### 3.6.3 Drzewa decyzyjne

Ostatnim rozważanym w tej pracy mechanizmem klasyfikacji są drzewa decyzyjne. Metoda ta wykorzystuje ideę konstrukcji drzewa do dzielenia ob-



Rys. 11: Dokładność modelu opartego o metodę drzew decyzyjnych w zależności od liczby iteracji boostingu

		Obserwacje		
		Blind	Other	Traffic
Predykcje	Blind	8	3	72
	Other	3	10	82
	Traffic	6	17	265

Tab. 15: Macierz błędów modelu opartego na metodzie drzew decyzyjnych

serwacji na bardziej jednorodnej grupie. W przygotowanym modelu do podziału drzewa w ramach węzła decyzyjnego był wykorzystywany algorytm C5.0, którego gotowa do wykorzystania implementacja jest dystrybuowana w środowisku R w ramach pakietu Caret.

Podobnie jak w przypadku poprzednich metod, tak i tym razem przeprowadzono proces wyboru najlepszych hiperparametrów modelu. Na wykresie (Rys. 11) zaprezentowane są zmiany dokładności modelu w zależności od przyjętej liczby iteracji boostingu.

Model osiąga najwyższą dokładność dla parametru określającego liczbę iteracji boostingu równego 8 i wynosi ona wtedy 0,61. Macierz błędów otrzymana dla zbioru testowego została zaprezentowana w tabeli (Tab. 15).

Wyniki działania modelu zbudowanego w oparciu o metodę drzew decyzyjnych wydają się być nieznacznie lepsze od dwóch pozostałych metod. Dokładność przygotowanych modeli w każdym z trzech rozważanych przy-

---

	„Blind”	„Other”	„Traffic”
Wrażliwość	0,47	0,33	0,63
Swoistość	0,83	0,80	0,51

Tab. 16: Wrażliwość i swoistość dla poszczególnych klas modelu opartego o metodę drzew decyzyjnych

padków jest właściwie taka sama. Na korzyść drzew decyzyjnych przemawia fakt bardziej zrównoważonego rozkładu wartości wrażliwości i swoistości dla poszczególnych klas (Tab. 16), co pozwala zakładać, że dla innych danych testowych, na przykład takich, w których stosunkowo mało przejazdów należy do klasy „Traffic”, model ten będzie działał lepiej od dwóch pozostałych.



# Rozdział 4

## Podsumowanie

Na zakończenie przedstawione zostanie zbiorcze podsumowanie przeprowadzonych prac oraz płynące z nich wnioski. Pokazane zostaną również potencjalne możliwości kontynuacji badań wykonanych w ramach niniejszej pracy.

### 4.1 Przeprowadzone prace

Patrząc całościowo, zasadnym wydaje się podział przeprowadzonych prac na dwa główne etapy. Pierwszy obejmował głównie wybór technologii i stworzenie odpowiednich narzędzi, które miały umożliwić przeprowadzenie analiz z wykorzystaniem dostępnych danych. W zakresie tego etapu było również czyszczenie i przygotowanie danych do dalszych badań. Głównym efektem pierwszego etapu było stworzenie aplikacji umożliwiającej modelowanie i analizę ruchu drogowego.

Etap drugi to z kolei intensywne wykorzystanie artefaktów etapu poprzedniego i skoncentrowanie się na badaniach oraz analizie dostępnych danych opisujących ruch drogowy, a także próby opracowania skutecznych modeli klasyfikacyjnych. Większość uwagi poświęcona została zagadnieniu, czy kierowcy poruszają się po trasach optymalnych oraz jak reagują na dynamicznie zmieniającą się sytuację na drogach i, czy uwzględniają informacje o ruchu drogowym podczas planowania trasy. Na zakończenie, przy wykorzystaniu danych z poprzednich eksperymentów, porównane zostały trzy modele klasyfikacyjne działające w oparciu o różne metody statystyczne.

## 4.2 Wnioski

Głównym celem pracy było stworzenie, na podstawie dostępnych danych historycznych, narzędzia modelującego ruch drogowy w sposób jak najbliższy temu, co jest obserwowane w rzeczywistości. Pierwszym wnioskiem nasuwającym się po wykonaniu wszystkich zaplanowanych zadań, ale również i po zapoznaniu się z wybraną literaturą dotyczącą podobnych zagadnień, jest stwierdzenie, że analiza zachowań kierowców jest bardzo złożonym zagadnieniem, na które wpływ ma wiele różnych czynników.

W celu pewnego usystematyzowania prac, po przeprowadzeniu pierwszych wstępnych analiz, jako punkt odniesienia przyjęte zostało założenie, że kierowcy zazwyczaj starają poruszać się trasami optymalnymi. Taka teza implikowała z kolei zdefiniowanie co to znaczy, że trasa jest optymalna. W pracy zbadano pojęcie trasy optymalnej pod kątem jak najmniejszej długości przejazdu, a także jak najkrótszego czasu przejazdu. Z przeprowadzonych analiz wynika, że rzeczywiste trasy przejazdów wybierane przez kierowców bliższe są trasom optymalnym pod kątem minimalizacji czasu przejazdu pomiędzy zadanymi punktami.

W dalszej części prac kwestia optymalizacji czasu przejazdu została rozszerzona o pojęcie wpływu aktualnego natężenia ruchu drogowego na wyznaczanie trasy optymalnej. W ten sposób w badaniach porównywane były dwa warianty trasy optymalnej. Pierwsze zestawienie wyników pokazało, iż wariant trasy optymalnej uwzględniający bieżącą sytuację na drogach nieznacznie lepiej odwzorowuje rzeczywiste decyzje kierowców.

Porównując trasy optymalne w tych dwóch wariantach, interesującym wydawało się być znalezienie pewnych cech charakterystycznych dla przejazdów, które będą określały, iż w danych warunkach przejazd rzeczywisty będzie bliższy jednej bądź drugiej opcji. Przeanalizowane zostały takie aspekty jak długość trasy, pora dnia czy zysk wynikający z uwzględniania informacji o bieżącym ruchu drogowym. Sformułowana została hipoteza mówiąca, iż im więcej czasu pozwala zaoszczędzić uwzględnienie informacji o aktualnym natężeniu ruchu, tym chętniej kierowcy będą takie informacje wykorzystywali. Jej słuszność można zaobserwować w wynikach kilku przeprowadzonych eksperymentów.

Jedną z cech charakterystycznych analizowanych danych był okres z jakiego pochodziły dane z Warszawy. Były to dwa kolejne czwartki lutego 2015 roku, pomiędzy którymi miał miejsce pożar Mostu Łazienkowskiego, który spowodował jego długotrwałe zamknięcie dla ruchu drogowego. Taka charakterystyka danych sprawiła, że znaczna część prac dotyczyła analizy przejazdów wiodących przez warszawskie mosty. Dla tych przejazdów sformułowane zostało inne kryterium klasyfikujące kierowców jako uwzględniających dane



o aktualnym natężeniu ruchu drogowego bądź nie. Oparte było ono o most, po którym odbywał się przejazd. W kolejnym etapie tak przygotowane dane posłużyły do porównania wybranych statystycznych modeli klasyfikacyjnych. Przeprowadzone testy pozwoliły na sformułowanie tezy, iż model oparty na drzewach decyzyjnych klasyfikuje kierowców nieznacznie lepiej niż modele oparte na metodzie k najbliższych sąsiadów czy metodzie Naive Bayes.

Przeprowadzone prace pokazały, że bardzo ciężko jest znaleźć uniwersalny sposób modelowania tak złożonego zagadnienia jak ruch drogowy. Jednakże wykonane analizy i stworzone narzędzia mogą to zadanie ułatwić. Kombinacja ich możliwości, tak jak to miało miejsce podczas badań przeprowadzonych w ramach niniejszej pracy, może czynić je przydatnymi elementami ułatwiającymi symulację zmian natężenia ruchu drogowego w zadanych warunkach dla wybranych grup kierowców.

### 4.3 Kontynuacja prac

Wyniki oraz wnioski płynące z niniejszej pracy mogą posłużyć za podstawę do kolejnych prac badawczych z zakresu analizy ruchu drogowego. Pewnym problemem podczas przeprowadzania niektórych eksperymentów był stosunkowo niewielki rozmiar próbki danych. Naturalnym więc mogłoby być zwerifikowanie powyższych wniosków dla większej liczby przejazdów. Większy zbiór danych mógłby również posłużyć do rozszerzenia analizy o wyszukiwanie kolejnych wzorców i cech charakterystycznych dla poszczególnych grup przejazdów. Dodatkowo, przeprowadzone porównanie modeli klasyfikacyjnych mogłoby zostać rozbudowane o kolejne bardziej zaawansowane metody, bądź, nawet jeszcze bardziej ogólnie, można by wykorzystać metody nienadzorowanego uczenia maszynowego do szukania wzorców charakteryzujących kierowców poruszających się po drogach publicznych.



# Bibliografia

- [1] [https://pl.wikipedia.org/wiki/Global\\_Positioning\\_System](https://pl.wikipedia.org/wiki/Global_Positioning_System)
- [2] [https://pl.wikipedia.org/wiki/Paradoks\\_Braessa](https://pl.wikipedia.org/wiki/Paradoks_Braessa)
- [3] Dietrich Braess, *Über ein Paradoxon aus der Verkehrsplanung*. „*Unternehmensforschung*“, 12, s. 258–268, 1968
- [4] <https://www.nytimes.com/1990/12/25/health/what-if-they-closed-42d-street-and-nobody-noticed.html?scp=8&st=nyt>
- [5] Asad J. Khattak, Joseph L. Schofer, and Frank S. Koppelman, *Effect of Traffic Reports on Commuters' Route and Departure Time Changes*, Vehicle Navigation and Information Systems Conference, 1991
- [6] Asad J. Khattak, Andre de Palma, *The impact of adverse weather conditions on the propensity to change travel decisions: A survey of Brussels commuters*, *Tronspn Res.-A*, Vol. 31, No. 3, pp: 181-203., 1997
- [7] <https://www.theatlantic.com/technology/archive/2018/03/mapping-apps-and-the-price-of-anarchy/555551/>
- [8] Rong-Chang Jou, Soi-Hoi Lam, Mei-Ch uan Weng, Chih-Cheng Chen, *Real Time Traffic Information and Its Impacts on Route Switching Behavior of Expressway Drivers*, *Journal of Advanced Transportation*, Vol. 38, No. 2, p p 187-223, 2004
- [9] Antonio Lima, Rade Stanojevic, Dina Papagiannaki, Pablo Rodriguez, and Marta C. González, *Understanding individual routing behaviour*, *Journal of The Royal Society Interface*, Volume 13, Issue 116, 2016
- [10] <https://theconversation.com/recalculating-by-not-driving-the-optimal-route-youre-causing-traffic-jams-56135>

- [11] Karthik K Srinivasan, Hani S. Mahmassani, *Modeling Inertia and Compliance Mechanisms in Route Choice Behavior Under Real-Time Information*, Transportation Research Record Journal of the Transportation Research Board, 2000
- [12] Grantham K. H. Pang, Kazunori Takahashi, Takayoshi Yokota, Hiroshi Takenaga, *Adaptive Route Selection for Dynamic Route Guidance System Based on Fuzzy-Neural Approaches*, IEEE Transactions on Vehicular Technology, vol. 48, no. 6, 1999
- [13] Corrado de Fabritiis, Roberto Ragona, Gaetano Valenti, *Traffic Estimation And Prediction Based On Real Time Floating Car Data*, Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems Beijing, China, October 12-15, 2008
- [14] Wanli Min, Laura Wynter, Yasuo Amemiya, *Road Traffic Prediction with Spatio-Temporal Correlations*, IBM Research Report RC24275, 2007
- [15] <https://yanosik.pl/>
- [16] <https://www.mysql.com/>
- [17] <https://qgis.org/en/site/>
- [18] <https://www.graphhopper.com/>
- [19] <https://www.r-project.org/>
- [20] Max Kuhn, *Package ‘caret’, version 6.0-83, Classification and Regression Training*, 2019 (<https://cran.r-project.org/web/packages/caret/caret.pdf>)
- [21] E. W. Dijkstra, *A note on two problems in connexion with graphs*, Numerische Mathematik, strony: 269–271, 1959
- [22] [https://pl.wikipedia.org/wiki/Algorytm\\_A\\*](https://pl.wikipedia.org/wiki/Algorytm_A*)
- [23] [https://en.wikipedia.org/wiki/Sinusoidal\\_projection](https://en.wikipedia.org/wiki/Sinusoidal_projection)
- [24] *Ustawa z dnia 20 czerwca 1997 r. Prawo o ruchu drogowym, Art. 2. pkt. 29 i 30*
- [25] [https://pl.wikipedia.org/wiki/Most\\_%C5%81azienkowski](https://pl.wikipedia.org/wiki/Most_%C5%81azienkowski)
- [26] [https://en.wikipedia.org/wiki/K-nearest\\_neighbors\\_algorithm](https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm)

**Dodatek A**

**Szczegółowe wyniki**

Data	Godzina	Mediana w gr. B	Mediana w gr. T	Liczba przejazdów w gr. B	Liczba przejazdów w gr. T
13.02.2014	8:00 – 8:15	4,85	24,99	43	46
	8:15 – 8:30	7,95	24,79	36	58
	16:45 – 17:00	3,49	19,47	48	39
	17:00 – 17:15	1,62	23,77	44	60
	17:15 – 17:30	9,86	16,47	28	45
14.02.2014	8:00 – 8:15	1,82	21,59	51	46
	8:15 – 8:30	9,02	22,58	41	45
	16:45 – 17:00	2,51	25,43	47	44
	17:00 – 17:15	3,10	23,79	51	55
	17:15 – 17:30	1,56	23,95	38	37
15.02.2014	8:00 – 8:15	4,31	13,65	15	18
	8:15 – 8:30	6,16	25,14	8	20
	16:45 – 17:00	1,76	26,21	17	20
	17:00 – 17:15	9,37	23,23	15	20
	17:15 – 17:30	10,67	21,64	23	23
Średnia ważona		4,63	22,69		

Tab. 17: Mediana oszczędności czasu (w sekundach) z poruszania się trasą  $P_{T_i}$  w porównaniu do poruszania się trasą  $P_{B_i}$  dla danych z Poznania  
*gr. B* – kierowcy, którzy nie uwzględniają informacji o natężeniu ruchu drogowego  
*gr. T* – kierowcy, którzy uwzględniają informacje o natężeniu ruchu drogowego

Data	Godzina	Średnia w gr. B	Średnia w gr. T	Liczba przejazdów w gr. B	Liczba przejazdów w gr. T
13.02.2014	8:00 – 8:15	22,52	38,01	43	46
	8:15 – 8:30	29,87	46,98	36	58
	16:45 – 17:00	12,92	34,43	48	39
	17:00 – 17:15	13,24	46,99	44	60
	17:15 – 17:30	17,62	39,81	28	45
14.02.2014	8:00 – 8:15	19,25	41,90	51	46
	8:15 – 8:30	24,19	45,94	41	45
	16:45 – 17:00	11,99	44,31	47	44
	17:00 – 17:15	8,75	44,58	51	55
	17:15 – 17:30	17,22	49,92	38	37
15.02.2014	8:00 – 8:15	7,93	34,38	15	18
	8:15 – 8:30	18,32	36,71	8	20
	16:45 – 17:00	15,32	34,12	17	20
	17:00 – 17:15	23,77	45,37	15	20
	17:15 – 17:30	17,51	43,65	23	23
Średnia ważona		17,15	42,74		

Tab. 18: Średnia arytmetyczna oszczędności czasu (w sekundach) z poruszania się trasą  $P_{T_i}$  w porównaniu do poruszania się trasą  $P_{B_i}$  dla danych z Poznania

*gr. B* – kierowcy, którzy nie uwzględniają informacji o natężeniu ruchu drogowego

*gr. T* – kierowcy, którzy uwzględniają informacje o natężeniu ruchu drogowego

Data	Godzina	Śr. zysk w p. proc. w gr. B	Śr. zysk w p. proc. w gr. T	Liczba przejaz- dów w gr. B	Liczba przejaz- dów w gr. T
13.02.2014	8:00 – 8:15	3,99	9,24	43	46
	8:15 – 8:30	5,02	8,9	36	58
	16:45 – 17:00	2,79	7,75	48	39
	17:00 – 17:15	2,24	8,74	44	60
	17:15 – 17:30	3,63	7,98	28	45
14.02.2014	8:00 – 8:15	3,63	9,23	51	46
	8:15 – 8:30	4,02	8,75	41	45
	16:45 – 17:00	2,47	10,54	47	44
	17:00 – 17:15	2,11	8,05	51	55
	17:15 – 17:30	2,34	8,95	38	37
15.02.2014	8:00 – 8:15	1,19	4,84	15	18
	8:15 – 8:30	2,15	6,36	8	20
	16:45 – 17:00	2,54	5,22	17	20
	17:00 – 17:15	2,97	6,26	15	20
	17:15 – 17:30	2,4	7,94	23	23
Średnia ważona		3,02	8,35		

Tab. 19: Stosunek zysku czasu z poruszania się trasą  $P_{T_i}$  w porównaniu do poruszania się trasą  $P_{B_i}$  do czasu długości trwania przejazdu wyrażony w punktach procentowych dla danych z Poznania

*gr. B* – kierowcy, którzy nie uwzględniają informacji o natężeniu ruchu drogowego

*gr. T* – kierowcy, którzy uwzględniają informacje o natężeniu ruchu drogowego



Data	Godzina	Mediana w gr. B	Mediana w gr. T	Liczba przejazdów w gr. B	Liczba przejazdów w gr. T
12.02.2015	8:00 – 8:15	7,11	27,27	102	153
	8:15 – 8:30	7,72	23,98	96	121
	16:45 – 17:00	5,50	20,29	126	114
	17:00 – 17:15	9,12	25,13	114	98
	17:15 – 17:30	5,67	19,36	107	119
19.02.2015	8:00 – 8:15	6,93	25,53	133	158
	8:15 – 8:30	8,81	23,10	122	149
	16:45 – 17:00	7,15	26,65	114	122
	17:00 – 17:15	8,93	25,06	109	127
	17:15 – 17:30	8,11	20,83	125	153
Średnia ważona		7,49	23,78		

Tab. 20: Mediana oszczędności czasu (w sekundach) z poruszania się trasą  $P_{T_i}$  w porównaniu do poruszania się trasą  $P_{B_i}$  dla danych z Warszawy  
*gr. B* – kierowcy, którzy nie uwzględniają informacji o natężeniu ruchu drogowego  
*gr. T* – kierowcy, którzy uwzględniają informacje o natężeniu ruchu drogowego

Data	Godzina	Średnia w gr. B	Średnia w gr. T	Liczba przejazdów w gr. B	Liczba przejazdów w gr. T
12.02.2015	8:00 – 8:15	13,31	34,67	102	153
	8:15 – 8:30	14,32	37,47	96	121
	16:45 – 17:00	10,83	30,41	126	114
	17:00 – 17:15	14,29	38,85	114	98
	17:15 – 17:30	10,04	38,13	107	119
19.02.2015	8:00 – 8:15	13,56	37,04	133	158
	8:15 – 8:30	13,90	38,94	122	149
	16:45 – 17:00	15,48	37,17	114	122
	17:00 – 17:15	17,49	39,29	109	127
	17:15 – 17:30	14,33	42,54	125	153
Średnia ważona		13,73	37,55		

Tab. 21: Średnia arytmetyczna oszczędności czasu (w sekundach) z poruszania się trasą  $P_{T_i}$  w porównaniu do poruszania się trasą  $P_{B_i}$  dla danych z Warszawy

*gr. B* – kierowcy, którzy nie uwzględniają informacji o natężeniu ruchu drogowego

*gr. T* – kierowcy, którzy uwzględniają informacje o natężeniu ruchu drogowego

Data	Godzina	Śr. zysk w p. proc. w gr. B	Śr. zysk w p. proc. w gr. T	Liczba przejaz- dów w gr. B	Liczba przejaz- dów w gr. T
12.02.2015	8:00 – 8:15	3,79	8,94	102	153
	8:15 – 8:30	4,54	9,12	96	121
	16:45 – 17:00	3,4	8,6	126	114
	17:00 – 17:15	4,49	9,76	114	98
	17:15 – 17:30	3,71	9,54	107	119
19.02.2015	8:00 – 8:15	4,14	9,28	133	158
	8:15 – 8:30	4,83	10,19	122	149
	16:45 – 17:00	5,37	10,39	114	122
	17:00 – 17:15	5,1	9,63	109	127
	17:15 – 17:30	4,76	11,4	125	153
Średnia ważona		4,41	9,71		

Tab. 22: Stosunek zysku czasu z poruszania się trasą  $P_{T_i}$  w porównaniu do poruszania się trasą  $P_{B_i}$  do czasu długości trwania przejazdu wyrażony w punktach procentowych dla danych z Warszawy

*gr. B* – kierowcy, którzy nie uwzględniają informacji o natężeniu ruchu drogowego

*gr. T* – kierowcy, którzy uwzględniają informacje o natężeniu ruchu drogowego

	Zysk z poruszania się nową trasą optymalną [w sekundach]	„Blind”	„Traffic”	„Other”	Odsetek kierowców uwzględniający informacje o ruchu drogowym	Odsetek kierowców uwzględniający informacje o ruchu drogowym (po odrzuceniu kierowców niekorzystających z nawigacji)
<b>12.02.2015</b>	(0; 25]	58	122	36	56%	68%
	(25; 50]	40	141	27	68%	78%
	(50; 100]	29	165	48	68%	85%
	(100; 150]	1	51	20	71%	98%
	Powyżej 150	0	20	13	61%	100%
<b>19.02.2015</b>	(0; 25]	58	96	35	51%	62%
	(25; 50]	37	129	41	62%	78%
	(50; 100]	37	154	64	60%	81%
	(100; 150]	13	65	23	64%	83%
	Powyżej 150	2	20	10	63%	91%
<b>Razem</b>	(0; 25]	116	218	71	54%	65%
	(25; 50]	77	270	68	65%	78%
	(50; 100]	66	319	112	64%	83%
	(100; 150]	14	116	43	67%	89%
	Powyżej 150	2	40	23	62%	95%

Tab. 23: Szczegółowe rezultaty pokazujące jaka część kierowców uwzględniła informacje o bieżącym natężeniu ruchu drogowego w zależności od stopnia w jakim uwzględnienie takiej informacji skraca czas przejazdu

	Długość trasy „Blind” [km]	Długość trasy „Traffic” [km]	Czas przejazdu [s]	A  bridge_traffic – bridge_real	B  bridge_blind – bridge_real	C  bridge_traffic – bridge_blind
Długość trasy „Blind” [km]	1	0,99	0,50	0,12	0,14	0,08
Długość trasy „Traffic” [km]	0,99	1	0,52	0,15	0,18	0,14
Czas przejazdu [s]	0,50	0,52	1	0,26	0,29	0,20
A  bridge_traffic – bridge_real	0,12	0,15	0,26	1	0,72	0,22
B  bridge_blind – bridge_real	0,14	0,18	0,29	0,72	1	0,01
C  bridge_traffic – bridge_blind	0,08	0,14	0,20	0,22	0,53	1

Tab. 24: Macierz korelacji