

WOJSKOWA AKADEMIA TECHNICZNA

im. Jarosława Dąbrowskiego



ROZPRAWA DOKTORSKA

BADANIE EMISJI CO₂ POJAZDU OSOBOWEGO NA WYBRANYM PRZYKŁADZIE Z ZASTOSOWANIEM METOD MATEMATYCZNYCH

Autor: mgr inż. Magdalena Rykała

Promotor: płk dr hab. inż. Anna Borucka, prof. WAT

Promotor pomocniczy: ppłk dr inż. Małgorzata Grzelak

STRESZCZENIE

Emisja gazów cieplarnianych, do których należy m.in. dwutlenek węgla (CO_2) dotyczy wielu obszarów gospodarki, a w szczególności sektora transportu. Dekarbonizacja transportu, oprócz zmian systemowych, wymaga także działań ukierunkowanych na pojedynczych użytkowników. Kształtowanie zachowania kierowców na drodze może mieć znaczny wpływ na emisję gazów cieplarnianych. Skuteczność indywidualnych strategii dążących do ograniczania emisji CO_2 w transporcie zależy od dostępności narzędzi, które wspomagają kierowców w zakresie ekologicznej jazdy i wskazuje na potrzebę ich tworzenia. Dlatego za cel naukowy pracy przyjęto opracowanie modelu oceny kształtowania poziomu emisji CO_2 na przykładzie wybranego pojazdu osobowego na poziomie pojedynczego uczestnika ruchu drogowego.

W ramach wprowadzenia do zagadnienia przedstawiono problematykę związaną z emisją CO_2 , wyjaśniono jej wpływ na środowisko naturalne, podkreślając obecne cele polityki klimatycznej kształtowanej przez UE. Wytypowano także trzy główne obszary dotyczące badań w zakresie emisji CO_2 w kontekście pojazdów samochodowych z silnikami spalinowymi: styl jazdy kierowców, zużycie paliwa oraz możliwości wykorzystania danych z systemów diagnostycznych. Następnie opracowano i opisano metodykę badań. Sformułowano cel główny rozprawy oraz cele szczegółowe, główny problem badawczy oraz problemy cząstkowe, a także postawiono tezę badawczą. Następnie wybrano kluczowe zmienne do badania oraz zaproponowano i scharakteryzowano wykorzystywane w rozprawie metody badawcze, opracowując algorytm planowanych działań. Próbkę badawczą stanowiły zebrane w toku badań eksperymentalnych obserwacje zawierające zapisy serii przejazdów testowych pojazdem osobowym wyposażonym w silnik spalinowy zasilany benzyną, pozyskane za pomocą interfejsu diagnostycznego OBD. Do badania wyodrębniono trzy podzbiory związane z pracą silnika: podczas ruszania pojazdu i jazdy z małymi prędkościami (tryb ruszania pojazdu), w trybie jazdy miejskiej oraz w trybie jazdy pozamiejskiej. Dla analizowanych podzbiorów skonstruowano modele regresji wieloczynnikowej oraz modele drzew decyzyjnych, wykorzystując algorytm CART, model lasu losowego oraz drzewa wzmacniane w celu analizy wpływu wytypowanych czynników na emisję CO_2 . Otrzymane modele poddano szczegółowej ocenie i wzajemnemu porównaniu, uwzględniając ich skuteczność predykcyjną oraz interpretowalność wyników. Przeprowadzona analiza pozwoliła na identyfikację zmiennych o najsilniejszym wpływie na emisję CO_2 oraz sformułowanie końcowych wniosków dotyczących kluczowych czynników determinujących poziom emisji. Wyniki badania dostarczają istotnych informacji umożliwiających optymalizację strategii redukcji emisji z punktu widzenia pojedynczego użytkownika oraz wskazują na najbardziej efektywne kierunki działań w zakresie polityki środowiskowej i technologicznych rozwiązań ograniczających emisję gazów cieplarnianych.

Słowa kluczowe: emisja CO_2 z pojazdu osobowego, modelowanie matematyczne, drzewa decyzyjne, regresja wieloczynnikowa

A study of the CO₂ emissions of a passenger vehicle on a selected example using mathematical methods

ABSTRACT

Emissions of greenhouse gases, which include carbon dioxide (CO₂), affect many areas of the economy, especially the transportation sector. The decarbonization of transportation, in addition to systemic changes, also requires measures aimed at individual users. Shaping driver behavior on the road can have a significant impact on greenhouse gas emissions. The effectiveness of individual strategies seeking to reduce CO₂ emissions in transportation depends on the availability of tools that assist drivers in green driving, and points to the need to create them. Therefore, the development of a model for evaluating the formation of CO₂ emissions on the example of a selected passenger vehicle at the level of the individual road participant was adopted as the scientific goal of the study.

As part of the introduction to the issue, the problem of CO₂ emissions is presented, its impact on the environment is explained, highlighting the current climate policy goals shaped by the EU. Three main areas for research on CO₂ emissions in the context of internal combustion engine vehicles were also identified: drivers' driving style, fuel consumption and the possibility of using data from diagnostic systems. The research methodology was then developed and described. The main objective of the dissertation was formulated, as well as specific objectives, the main research problem and sub-problems, and the research hypothesis was set. Then the key variables for the study were selected, and the research methods used in the dissertation were proposed and characterized, developing an algorithm for the planned activities. The research sample consisted of observations collected in the course of the experimental research, which included records of a series of test runs on a passenger vehicle equipped with a gasoline-powered internal combustion engine, acquired using the OBD diagnostic interface. Three subsets related to engine operation were extracted for the study: during vehicle start-up and low-speed driving (vehicle start-up mode), in urban driving mode and during driving in non-urban mode. Multivariate regression models and decision tree models were constructed for the subsets analyzed, using the CART algorithm, the random forest model and boosted trees to analyze the effects of the selected factors on CO₂ emissions. The obtained models were subjected to detailed evaluation and peer comparison, taking into account their predictive effectiveness and interpretability of the results. The analysis made it possible to identify the variables with the strongest impact on CO₂ emissions and to formulate final conclusions on the key factors determining the level of emissions. The results of the study provide important information for optimizing emission reduction strategies from the point of view of a single user, and indicate the most effective courses of action in environmental policy and technological solutions for reducing greenhouse gas emissions.

Keywords: CO₂ emissions of a passenger vehicle, mathematical modeling, decision trees, multivariate regression

SPIS TREŚCI

WYKAZ SKRÓTÓW	6
WYKAZ OZNACZEŃ	8
1. WSTĘP	9
2. UZASADNIENIE TEMATU BADAŃ	13
2.1. Emisyjność CO ₂ z transportu	13
2.2. Przegląd metod modelowania emisyjności pojazdów samochodowych	34
2.3. Styl jazdy kierowcy w ujęciu literaturowym	39
2.4. Zużycie paliwa a emisja CO ₂	48
2.5. Badanie systemów diagnostycznych w pojazdach samochodowych	52
3. CEL PRACY I ZAŁOŻENIA BADAWCZE	61
4. METODY MODELOWANIA MATEMATYCZNEGO ZASTOSOWANE W BADANIU	68
4.1. Teoretyczne aspekty modelowania	68
4.2. Algorytm badania oraz wstępna analiza danych	74
4.2.1. Charakterystyka zmiennych wykorzystanych w badaniu	74
4.2.2. Modelowanie za pomocą metody regresji wielorakiej	89
4.2.3. Modelowanie z wykorzystaniem metody drzew decyzyjnych	97
5. MODELE EMISJI CO ₂ DO ŚRODOWISKA	107
5.1. Wstępna analiza danych oraz przebiegu zmiennej objaśnianej i objaśniającej	107
5.2. Modelowanie emisji CO ₂ z zastosowaniem modelu regresji wielorakiej	115
5.2.1. Model regresji wieloczynnikowej emisji CO ₂ dla ruszania pojazdu	115
5.2.2. Model regresji wieloczynnikowej w trybie jazdy miejskiej	121
5.2.3. Model regresji wieloczynnikowej w trybie jazdy pozamiejskiej	126
5.2.4. Porównanie modeli regresji wieloczynnikowej w trzech trybach jazdy	131
5.3. Modelowanie emisji CO ₂ z zastosowaniem modeli drzew decyzyjnych	133
5.3.1. Modele regresyjnych drzew decyzyjnych dla momentu ruszania pojazdu	133
5.3.2. Modele regresyjnych drzew decyzyjnych dla jazdy w trybie miejskim	137
5.3.3. Modele regresyjnych drzew decyzyjnych podczas jazdy w trybie pozamiejskim	141
5.3.4. Porównanie modeli w poszczególnych trybach jazdy	145
6. ZAKOŃCZENIE	149
6.1. Podsumowanie przeprowadzonych badań	149
6.2. Wnioski	152
LITERATURA	156

WYKAZ SKRÓTÓW

ABS	(z ang. Anti-Lock Braking System) Układ zapobiegający blokowaniu kół podczas hamowania
ACEA	(z ang. European Automobile Manufacturers' Association) Europejskie Stowarzyszenie Producentów Pojazdów
AIC	(z ang. Akaike Information Criterion) Kryterium informacyjne Akaikego
ANN	(z ang. Artificial Neural Networks) Sztuczne sieci neuronowe
ASR	(z ang. Acceleration Slip Regulation) Układ kontroli trakcji
BEV	(z ang. Battery Electric Vehicle) Samochód elektryczny (akumulatorowy)
CART	(z ang. Classification and Regression Trees) Algorytm regresji i klasyfikacji
CHTC-HT	(z ang. China Automotive Test Cycle of Heavy-Duty Trucks) Cykl testowy ciężkich pojazdów użytkowych w Chinach
CNN	(z ang. Convolutional Neural Network) Głębokie sieci konwolucyjne
DLC	(z ang. Data Link Connector) Złącze łączy danych
ECU	(z ang. Engine Control Unit) Elektroniczny moduł sterujący
EEA	(z ang. European Environment Agency) Europejska Agencja Środowiska
EOBD	(z ang. Euro-On-Board-Diagnostic) Europejski system diagnostyki pokładowej
ESP	(z ang. Electronic Stability Program) Układ stabilizacji toru jazdy
GHG	(z ang. Greenhouse Gases) Gazy cieplarniane
GNSS	(z ang. Global Navigation Satellite Systems) Globalny System Nawigacji Satelitarnej
GWP	(z ang. Global Warming Potential) Potencjał tworzenia efektu cieplarnianego
HEV	(z ang. Hybrid Electric Vehicle) Samochód hybrydowy
HFC	(z ang. Fluorocarbons) Fluorowęglowodory
IMU	(z ang. Inertial Measuring Unit) Bezwładnościowa jednostka pomiarowa
INS	(z ang. Inertial Navigation System) Układ nawigacji bezwładnościowej
k-NN	(z ang. K Nearest Neighbours) Metoda k-najbliższych Sąsiadów
LDV	(z ang. Light Duty Vehicle) Lekki pojazd samochodowy
LSTM	(z ang. Long Short-Term Memory) Rodzaj rekurencyjnej sieci neuronowej
MIL	(z ang. Malfunction Indicator Light) Kontrolka sygnalizująca usterki
MTF	(z ang. Mean Tractive Force) Średnia siła ciągu
OBD	(z ang. On-Board Diagnostics) System pokładowej diagnostyki
MSE	(z ang. Mean Squared Error) Średni błąd kwadratowy
MAE	(z ang. Mean Absolute Error) Średni błąd bezwzględny
MRAE	(z ang. Mean Relative Absolute Error) Średni względny błąd bezwzględny
MAPE	(z ang. Mean Absolute Percentage Error) Średni bezwzględny błąd procentowy
PEMS	(z ang. Portable Emissions Measurement Systems) Mobilny system pomiaru emisji
PFC	(z ang. Perfluorocarbons) Perfluorowęglowodory
PHEV	(z ang. Plug-in Hybrid Electric Vehicle) Samochód hybrydowy z możliwością ładowania ze źródła zewnętrznego

WYKAZ SKRÓTÓW – ciąg dalszy

PKB	(z ang. Gross Domestic Product, GDP) Produkt Krajowy Brutto
PMML	(z ang. Predictive Model Markup Language) Format XML służący do opisu modeli eksploracji danych i modeli statystycznych
SST	(z ang. Sum of Squares Total) Zmienność całkowita
SSW	(z ang. Sum of Squares Within Groups) Zmienność wewnątrzgrupowa
SSB	(z ang. Sum of Squares Between Groups) Zmienność międzygrupowa
SVM	(z ang. Support Vector Machines) Metoda wektorów nośnych
TPMS	(z ang. Tire Pressure Monitoring System) System monitorujący ciśnienie w oponach
UE	(z ang. European Union, EU) Unia Europejska
VIN	(z ang. Vehicle Identification Number) Numer identyfikacyjny pojazdu
ZI	(z ang. Petrol Engine) Silnik o zapłonie iskrowym
ZS	(z ang. Diesel Engine) Silnik o zapłonie samoczynnym

WYKAZ OZNACZEŃ

n/N	prędkość obrotowa silnika / liczba obserwacji
Δr	przyrost liczby obrotów wału silnika
Δt	przyrost czasu
$L_{\%}$	względne obciążenie silnika
P / P_{max}	moc silnika / maksymalna moc silnika
W	praca mechaniczna
M	moment obrotowy
V	prędkość liniowa
Δs	przyrost drogi
F_T	siła styczna
r	promień wykorbienia wału korbowego
$\varphi_{\%}$	względne położenie przepustnicy
φ / φ_{max}	kąt otwarcia przepustnicy / maksymalny kąt otwarcia przepustnicy
$\dot{m}_p$	natężenie przepływu paliwa
q	masa wtryskiwanego paliwa do cylindra na jeden cykl
z	liczba cylindrów silnika
τ	liczba obrotów wału korbowego silnika na cykl
m_{CO_2}	wskaźnik emisji CO ₂
k_{CO_2}	współczynnik emisji CO ₂ dla paliwa kopalnego
X	wektor obserwacji $x=(x_1, x_2, \dots, x_n)$
$\bar{x}$	średnia arytmetyczna zmiennej x
Me	mediana
σ	odchylenie standardowe
K	kurtoza
A	skośność
μ	moment centralny
F_n	dystrybuanta empiryczna
F	dystrybuanta teoretyczna
D_n	statystyka testu
J	funkcja celu
H	hipoteza testowa
C	centrum zbioru
Y	wartość empiryczna zmiennej zależnej y
x	wartość zmiennej niezależnej x
$\hat{y}_t$	wartość teoretyczna wyznaczone na podstawie modelu
β	parametr modelu regresji
ε	składnik losowy
R^2	współczynnik determinacji modelu
s_d	odchylenie standardowe
r_{xy}	współczynnik korelacji Pearsona
$cov(x, y)$	kowariancja zmiennych x oraz y
I	współczynnik Ginniego
E	entropia
ε_g	błąd generalizacji
p	prawdopodobieństwo wystąpienia zdarzenia

1. WSTĘP

Redukcja emisji substancji szkodliwych pochodzących z pojazdów samochodowych stanowi obecnie jedno z najważniejszych obserwowanych w Unii Europejskiej (UE) wyzwań środowiskowych, zobowiązując kraje członkowskie do zmian w tym zakresie. Wymagania dotyczą emisji gazów cieplarnianych, do których należą m.in. dwutlenek węgla (CO_2), metan (CH_4), podtlenek azotu (N_2O). Transport uznawany jest przez Unię Europejską za jeden z sektorów gospodarki o najbardziej szkodliwym wpływie na środowisko naturalne i zdrowie ludzi, gdyż odgrywa znaczną rolę w kształtowaniu poziomu emisji CO_2 w Unii Europejskiej, odpowiadając za około 30% całkowitej emisji CO_2 . Około 70% tej wartości pochodzi z transportu drogowego, z czego aż 61% generują samochody osobowe [59].

Z tego powodu UE wprowadza regulacje, których celem jest dążenie do redukcji tej emisji. Polityka UE w tym obszarze polega m.in. na wprowadzaniu nowych, coraz bardziej restrykcyjnych norm emisji CO_2 dla nowych samochodów osobowych. Docelowo normy te mają na celu osiągnięcie neutralności klimatycznej do 2050 roku i redukcję emisji z transportu o 90%. Działania te mają charakter globalny, jednak pewne działania lokalne mogą również mieć wpływ na redukcję emisji CO_2 . Jednym z kluczowych działań jest zwiększanie świadomości kierowców w zakresie zasad ekologicznej jazdy. Skuteczna dekarbonizacja transportu wymaga nie tylko zmian systemowych, ale także działań na poziomie operacyjnym, ukierunkowanych na pojedynczych użytkowników. Indywidualne nawyki kierowców, takie jak płynna jazda, unikanie nadmiernego przyspieszania i hamowania czy racjonalne wykorzystanie paliwa, mogą mieć istotny wpływ na całkowitą emisję gazów cieplarnianych. Aby strategie edukacyjne i motywacyjne na poziomie jednostkowym mogły przynieść wymierne efekty w skali globalnej, konieczne jest zapewnienie kierowcom odpowiednich narzędzi wspierających ich decyzje. Jednym z takich narzędzi mogą być modele matematyczne, opracowane na podstawie danych z systemu pokładowej diagnostyki silnika OBD (On-Board Diagnostic System), które precyzyjnie analizują wpływ różnych czynników, takich jak styl jazdy, obciążenie pojazdu czy warunki drogowe, na emisję CO_2 . Takie założenie stało się genezą niniejszej pracy. Działania tego typu są szczególnie ważne biorąc pod uwagę ciągle zwiększającą się liczbę pojazdów w UE. Z związku z powyższym w niniejszej rozprawie podjęto badania mające na celu opracowanie modelu oceny kształtowania poziomu emisji CO_2 na przykładzie wybranego pojazdu na poziomie pojedynczego uczestnika ruchu drogowego.

Rozważanie w tym zakresie rozpoczęto wstępem, w ramach którego przybliżono i scharakteryzowano omawianą w rozprawie tematykę. Następnie w rozdziale drugim pracy

przeprowadzono analizę stanu zagadnienia, w której skupiono się na przedstawieniu problematyki związanej z emisją CO₂ w odniesieniu do dostępnych badań w tym zakresie. Omówiono rodzaje gazów cieplarnianych, wyjaśniono ich wpływ na środowisko naturalne i przedstawiono ich znaczenie w kontekście postępujących zmian klimatu. Dodatkowo opisano w jaki sposób tematyka związana z gazami cieplarnianymi, a szczególnie z CO₂ wpisuje się w obecne cele polityki klimatycznej kształtowanej przez UE. Określono aktualny stan spełnienia norm emisji CO₂ w krajach UE ze szczególnym uwzględnieniem Polski. Wyszczególniono związane z tym problemy wykorzystując analizę danych statystycznych dotyczących emisji CO₂ w krajach UE. W ten sposób przedstawiono szczegółowo tło dotyczące badanej problematyki emisji CO₂. W ramach tego rozdziału wytypowano także obszary tematyczne dotyczące emisji CO₂ w kontekście pojazdów samochodowych z silnikami spalinowymi. Na podstawie przeprowadzonej analizy literatury stwierdzono, że najbardziej znaczące czynniki w kontekście modelowania emisyjności pojazdów samochodach stanowią trzy obszary tematyczne: styl jazdy kierowców, zużycie paliwa oraz systemy diagnostyczne.

Analiza stanu zagadnienia w kontekście stylu jazdy wykazała, iż kierowca swoim zachowaniem na drodze może znacząco przyczynić się do redukcji śladu węglowego związanego z transportem drogowym. Kluczowymi aspektami w tym obszarze są prędkość pojazdu oraz dynamika jazdy. Nadmierna prędkość (agresywny styl jazdy) wiąże się najczęściej z większym zużyciem paliwa, co prowadzi do zwiększonej emisji CO₂. Jazda z umiarkowaną prędkością dostosowaną do warunków jazdy (zrównoważony styl), zazwyczaj w zakresie średnich obrotów silnika, jest bardziej efektywna pod względem zużycia paliwa oraz emisji CO₂. W kontekście dynamiki jazdy, gwałtowne przyspieszanie i hamowanie mogą znacznie zwiększyć zużycie paliwa, natomiast łagodne i płynne przyspieszanie i stopniowe hamowanie pozwalają redukować emisję CO₂.

Kolejnym obszarem tematycznym jest zużycie paliwa, które jest bezpośrednio powiązane z emisją CO₂. Analiza literatury w tym obszarze wykazała, iż istnieje szereg metod prognozowania stosowanych do modelowania zużycia paliwa, np. regresja liniowa, drzewa decyzyjne, sieci neuronowe. Bardzo ważnym zagadnieniem w tym obszarze jest określenie czynników związanych z pojazdem (parametry pracy silnika, parametry kinematyczne pojazdu, itd.) wpływających bezpośrednio na zużycie paliwa. W wyniku przeprowadzonej analizy literatury w tym zakresie wskazano, że istotny jest wpływ, m.in. prędkości pojazdu, prędkości obrotowej silnika, obciążenia silnika, mocy silnika itd. Skupiono się w tym przypadku na czynnikach dynamicznych (zmieniających się w czasie ruchu pojazdu), natomiast pominięto stałe parametry techniczne, tj. masa, kształt aerodynamiczny, opory toczenia kół itd.

Ostatnim wymienionym obszarem tematycznym są pokładowe systemy diagnostyki silnika. Systemy te umożliwiają monitorowanie emisji CO₂ oraz zużycia paliwa przez dostęp do wybranych parametrów diagnostycznych pojazdu. W pracy dokonano przeglądu literatury dotyczącego systemów pokładowej diagnostyki silnika (OBD). Na rynku istnieje wiele różnych odmian interfejsów diagnostycznych, wśród których należy wyróżnić szczególnie rozwiązania bezprzewodowe ze względu na ich uniwersalny charakter, możliwość integracji z systemami mobilnymi (smartfony) oraz dostępność.

W rozdziale trzecim przedstawiono cel i założenia badawcze pracy, tj. problem badawczy, problemy cząstkowe, cel główny rozprawy, cele szczegółowe oraz tezę badawczą. Na podstawie przeprowadzonej w rozdziale drugim analizy stanu zagadnienia określono parametry wejścia (prędkość pojazdu, prędkość obrotowa silnika, względne obciążenie silnika, moment obrotowy, moc silnika, względne położenie przepustnicy – dotyczy silnika ZI) i wyjścia (emisja CO₂) modelu. Pierwszym etapem badań były badania eksperymentalne, w których dokonano serii przejazdów testowych pojazdem osobowym wyposażonym w silnik spalinowy zasilany benzyną. Do zapisu wymienionych wcześniej zmiennych wykorzystano bezprzewodowy interfejs pomiarowy OBD połączony bezprzewodowo ze smartfonem. Testy przeprowadzono wykorzystując pojazd osobowy (Mazda 3, silnik 2.0, sedan). W czasie jazdy rejestrowano parametry pojazdu (wejścia oraz wyjście modelu), które umożliwiły stworzenie bazy danych stanowiącej podstawę późniejszych analiz. Trasy przejazdu były zróżnicowane pod kątem dozwolonej prędkości pojazdu, liczby zatrzymań, występowania odcinków dróg w obszarach miejskich i pozamiejskich.

Rozdział czwarty dotyczy metodyki przeprowadzonych badań, w którym na wstępie przedstawiono teoretyczne aspekty modelowania. Zostały one szczegółowo opisane z uwzględnieniem występujących w nich zjawisk fizycznych oraz stanowiących ich podstawę zależności matematycznych. Następnie na podstawie opisanych wejść i wyjść z modelu zaprezentowano opracowany algorytm planowanych badań wraz z wyszczególnieniem stosowanych metod badawczych. Rozdział kończy się opisem teoretycznym wybranych metod modelowania emisji CO₂ wykorzystanych w dysertacji, tj. regresji wieloczynnikowej oraz regresyjnych drzew decyzyjnych.

W rozdziale piątym skupiono się na modelowaniu emisji CO₂ z pojazdu osobowego do środowiska. Dokonano wstępnej analizy danych oraz analizy przebiegu zmiennych objaśniających i objaśnianej, gdzie na ich podstawie dokonano wyboru odpowiednich modeli matematycznych. Następnie zidentyfikowano grupy o zbliżonych charakterystykach oraz dokonano podziału obserwacji na jednorodne grupy obiektów, wykorzystując w tym celu

analizę skupień za pomocą metody k-średnich. Okazało się, że analizę przebiegu funkcji emisji CO₂ w zależności od wytypowanych parametrów należy podzielić na trzy grupy, związane z pracą silnika podczas ruszania pojazdu i jazdy z niskimi prędkościami (tryb ruszania pojazdu), w trybie jazdy miejskiej oraz podczas jazdy w trybie pozamiejskim. Dla zidentyfikowanych, jednorodnych grup w pierwszej kolejności skonstruowano modele regresji wieloczynnikowej, dla których dokonano oszacowania parametrów, a następnie dokonano ich oceny i porównania. Zaobserwowano, iż największy poziom emisji CO₂ do środowiska ma miejsce w trybie ruszania (365,60 g/km), natomiast najmniejszy w czasie poruszania się pojazdu w trybie pozamiejskim (166,18 g/km). Ponadto stwierdzono, że niezależnie od trybu jazdy, każdorazowy wzrost prędkości jazdy powoduje zmniejszenie emisji CO₂ do środowiska.

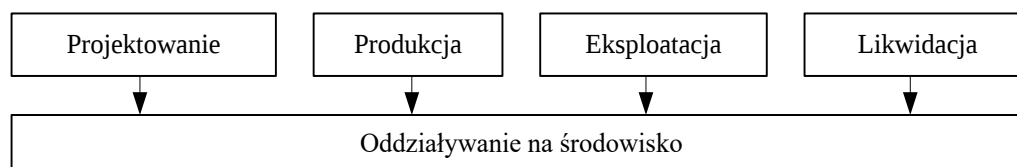
W kolejnej części omawianego rozdziału do badania wpływu wytypowanych czynników na emisję CO₂ zastosowano modele drzew decyzyjnych z wykorzystaniem algorytmu CART, model lasu losowego oraz drzew wzmacnianych. Dla zbudowanych modeli zaprezentowano również ranking ważności predyktorów, który umożliwia ocenę wpływu poszczególnych zmiennych objaśniających na badane zjawisko. Na ich podstawie zaobserwowano, iż największy wpływ na emisję CO₂ w trybie ruszania oraz w trybie miejskim ma prędkość pojazdu oraz względne obciążenie silnika. W trybie ruszania najmniejszy wpływ na emisję CO₂ ma z kolei względne obciążenie przepustnicy (model drzewa CART). W modelu lasu losowego ponownie największy wpływ miała prędkość pojazdu oraz względne położenie przepustnicy, najmniejszy zaś względne obciążenie silnika.

W rozdziale szóstym dokonano zakończenia przeprowadzonych badań, gdzie przedstawiono podsumowanie z przeprowadzonych badań oraz sformułowano wnioski końcowe. Dodatkowo określono również kierunki dalszych badań.

2. UZASADNIENIE TEMATU BADAŃ

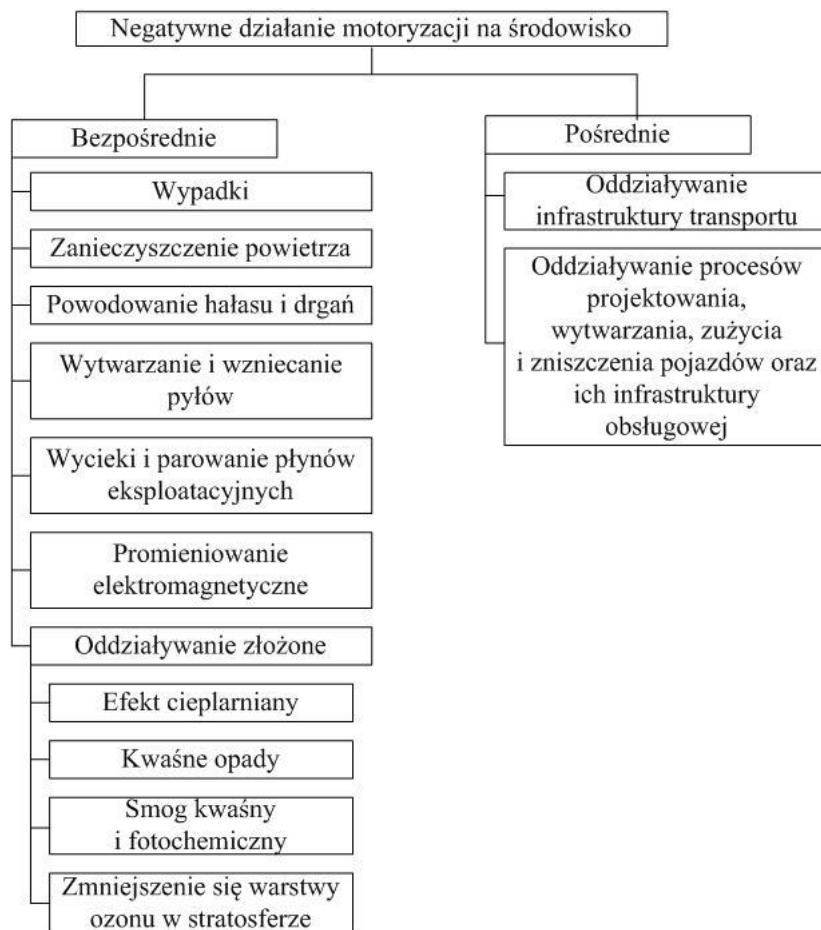
2.1. Emisyjność CO₂ z transportu

Motoryzacja jest obecnie nieodłącznym elementem życia codziennego społeczeństwa, który niewątpliwie napędza gospodarkę – przyspieszając jej rozwój, ale również zapewniając mobilność na niespotykaną skalę. Postęp technologiczny w działalności transportowej niesie ze sobą liczne korzyści, zarówno te społeczne jak i przede wszystkim ekonomiczne, jednakże jednocześnie powoduje wiele negatywnych skutków – oddziałując na środowisko naturalne. Aby zminimalizować negatywne oddziaływanie niezbędne jest podjęcie odpowiednich kroków zarówno globalnie, jak również jednostkowo (indywidualnie). Społeczeństwo obecnie staje się bardziej świadome i coraz większą wagę przywiązuje do oddziaływania motoryzacji na środowisko naturalne i życie człowieka [20], [41], [51]. Prowadzone badania dotyczące wpływu na środowisko wyróżniają podmioty i przedmioty oddziaływań. Do podmiotów oddziaływań możemy zaliczyć przede wszystkim: środki transportu, materiały eksploatacyjne pojazdów samochodowych, produkcję materiałów eksploatacyjnych, produkcję pojazdów oraz całą infrastrukturę drogową [19], [20]. Przedmiotami oddziaływań są wszystkie elementy środowiska, tj. teren, krajobraz, zasoby naturalne, powietrze, gleba, woda, rośliny, zwierzęta oraz ludzie [19], [20]. Wpływ motoryzacji występuje w każdej fazie eksploatacji urządzeń technicznych, zaczynając od fazy projektowania, a kończąc na fazie jej likwidacji (rys. 1).



Rys. 1. Wpływ poszczególnych faz motoryzacji na środowisko [19], [20], [128]

Na rysunku 2 przedstawiono podział szkodliwego oddziaływania na środowisko ze względu na jego bezpośredni i pośredni charakter. Do bezpośrednich skutków możemy zaliczyć: wypadki drogowe, zanieczyszczenie powietrza, wycieki i parowanie paliwa oraz innych płynów eksploatacyjnych itp. Natomiast do czynników pośrednich możemy zaliczyć: oddziaływanie infrastruktury transportu, oddziaływanie procesów projektowania, wytwarzania, zużycia i zniszczenia pojazdów, infrastruktury i bazy obsługowej transportu oraz oddziaływanie spowodowane czynnikami socjologicznymi i cywilizacyjnymi [20].



Rys. 2. Skutki oddziaływania motoryzacji na środowisko [65]

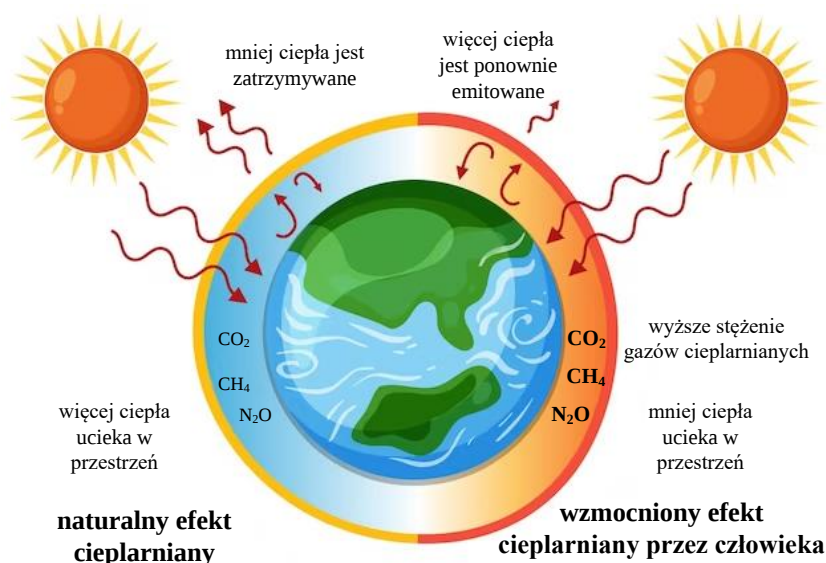
Poza podstawowymi skutkami oddziaływania motoryzacji na środowisko możemy również wyodrębnić działania złożone (rys. 2), obejmujące szereg różnorodnych czynników, które wspólnie oddziałują na ekosystemy, zasoby naturalne oraz przede wszystkim na jakość życia ludzkiego. Do najważniejszych oddziaływań możemy zaliczyć [41], [76], [128]:

- Emisję gazów cieplarnianych, która szczegółowo zostanie omówiona poniżej.
- Zanieczyszczenie powietrza – silniki spalinowe emitując szkodliwe substancje, mogą powodować problemy zdrowotne u społeczeństwa (tj. podrażnienia dróg oddechowych, problemy z oddychaniem, problemy sercowo-naczyniowe, podrażnienia skóry i oczu) oraz także negatywnie wpływać na faunę i florę. Ponadto mogą również powodować zanieczyszczenie powietrza w postaci [19]:
 - ✓ Smogu kwaśnego (londyńskiego) – który występuje głównie w obszarach miejskich i przemysłowych. Jest zjawiskiem, które charakteryzuje się występowaniem gęstych, trudnych do przewietrzenia mgieł zawierających duże stężenia dwutlenku siarki (SO_2) oraz tlenki azotu (NO_x). Substancje te emitowane są głównie przez spalanie paliw kopalnych w procesach przemysłowych i transportowych. W wyniku reakcji

z czynnikami atmosferycznymi, takimi jak wilgotność powietrza oraz promieniowanie słoneczne, tworzą aerozole i kwaśne deszcze [20].

- ✓ Kwaśnego deszczu - są to atmosferyczne opady o kwasowym odczynie pH niższym niż 5,6, spowodowane zanieczyszczeniem atmosferycznym przez tlenki siarki (głównie dwutlenek siarki SO_2) oraz tlenki azotu (głównie tlenek azotu NO_2), które reagują z wodą tworząc kwas siarkowy (H_2SO_4) i kwas azotowy (HNO_3) [21].
- ✓ Smogu fotochemicznego (kalifornijskiego) – który występuje głównie w obszarach miejskich o dużym natężeniu ruchu drogowego i przemysłowego, zwłaszcza w gorących i słonecznych warunkach. Jest zjawiskiem, w którym substancje chemiczne tj. tlenki azotu (NO_x) oraz lotne związki organiczne (VOC), w wyniku intensywnego promieniowania reagują ze sobą tworząc wysokie stężenie ozonu troposferycznego (O_3) oraz inne szkodliwe związki, takie jak aldehydy i azotan peroksyacetylu (PAN) [51].

Występujące w obecnych czasach postępujące zmiany klimatu są spowodowane narastającym efektem cieplarnianym (rys. 3). Do głównych przyczyn tego zjawiska można zaliczyć gazy cieplarniane GHG (z ang. Greenhouse Gases), które przepuszczają większość promieniowania słonecznego docierającego do Ziemi. Globalne ocieplenie i zmiany klimatyczne są spowodowane przez proces absorpcji i emisji promieniowania podczerwonego. Pochłaniając ciepło słoneczne, gazy cieplarniane tworzą tzw. „efekt szklarniowy” wokół Ziemi, prowadząc tym samym do zmian w rozkładzie opadów, wzrostu temperatur, czy też do fali upałów i powodzi. Gazy cieplarniane występują naturalnie w atmosferze, jednak działalność człowieka znacznie przyczynia się do ich akumulacji, co powoduje wzmocnienie efektu cieplarnianego.



Rys. 3. Zjawisko efektu cieplarnianego [95]

W celu porównywania wpływu różnych gazów na efekt cieplarniany i globalne ocieplenie utworzono pojęcie potencjału tworzenia efektu cieplarnianego GWP (z ang. Global Warming Potential), który jest miarą ich szkodliwości. Stanowi wskaźnik, który odnosi się do zdolności jednostki masy danego gazu do absorbowania ciepła w stosunku do tego samego wskaźnika dla dwutlenku węgla (CO_2) w określonym czasie, zazwyczaj dotyczącym setek lat.

Do głównych gazów cieplarnianych można zaliczyć:

- Dwutlenek węgla (CO_2) – jeden z głównych gazów cieplarnianych, który ma znaczący wpływ na globalny klimat. CO_2 jest najczęściej występującym gazem cieplarnianym w porównaniu do innych gazów. Według Europejskiej Agencji Środowiska (EEA) w 2019 roku dwutlenek węgla stanowił 80% emisji gazów cieplarnianych na świecie. Jest emitowany, m.in. przez spalanie paliw kopalnych, tj. węgiel, ropa naftowa, gaz ziemny oraz w wyniku realizacji istniejących procesów przemysłowych. Jego obecność w atmosferze możemy zaobserwować w postaci naturalnie występujących procesów, tj. oddychanie organizmów, erupcje wulkanów czy też procesy rozkładu materii organicznej, jak również w wyniku procesów antropogenicznych, tj. spalanie paliw kopalnianych, wycinanie lasów, użytkowanie ziemi itd. Dwutlenek węgla w naturalny sposób utrzymuje ciepło w atmosferze, co tym samym umożliwia istnienie życia na Ziemi, jednak wzrost stężenia CO_2 w atmosferze jest główną przyczyną zmian klimatycznych. Ponadto, jego wpływ na klimat jest długotrwały, ze względu na znaczny czas utrzymywania się w atmosferze (do tysiąca lat), w związku z czym stanowi, tzw. długotrwałe zanieczyszczenie atmosfery.
- Metan (CH_4) – jest drugim co do wielkości gazem cieplarnianym pod względem emisji na świecie (ok. 11%). Metan stanowi główny składnik gazu ziemnego oraz biometanu. Podobnie jak w przypadku dwutlenku węgla, obecność metanu w atmosferze ma zarówno charakter naturalny, jak i również powstający w wyniku działalności gospodarczej człowieka, tzn. wydobywanie ropy naftowej i węgla, wysypiska śmieci oraz rolnictwo. Dodatkowo, w porównaniu do dwutlenku węgla, stanowi on, tzw. krótkotrwałe zanieczyszczenie atmosfery (utrzymuje się w atmosferze ok. 10–15 lat), które posiada ok. 23-krotnie większą wartość potencjału tworzenia efektu cieplarnianego GWP, niż dwutlenek węgla. Ponadto, metan w atmosferze reaguje z aktywnymi związkami tlenu, utleniając się do dwutlenku węgla, co prowadzi do dalszego efektu cieplarnianego. Prowadzi również do zanieczyszczenia powietrza, co ma szkodliwy wpływ na cały ekosystem, w tym również ludzi.

- Podtlenek azotu (N_2O) – jest trzecim co do wielkości gazem cieplarnianym pod względem emisji na świecie (ok. 6%). Naturalnie w atmosferze występuje jedynie w niewielkich ilościach, a postępująca działalność rolnicza i przemysłowa znacznie zwiększyła jego stężenie. Głównymi źródłami emisji N_2O są wspomniana działalność rolnicza, procesy przemysłowe oraz spalanie paliw kopalnych i biomasy. Czas życia podtlenku azotu w atmosferze szacuje się na około 144 lata, natomiast jego wartość potencjału tworzenia efektu cieplarnianego GWP jest aż 296 większa od dwutlenku węgla.
- Fluorowane gazy cieplarniane – są czwartą grupą gazów cieplarnianych, które stanowią ok. 2% całkowitej emisji gazów cieplarnianych w Unii Europejskiej (UE). Spośród wszystkich gazów fluorowanych uznaje się, iż fluorowęglowodory (HFC), perfluorowęglowodory (PFC) i sześćiofluorek siarki (SF_6) mają największy wpływ na klimat. Wspólnymi cechami charakterystycznymi omawianych gazów jest to, iż wszystkie w swoim składzie zawierają fluor oraz wykazują wysoki współczynnik GWP w zakresie od 140 do 23000 w stosunku do CO_2 . Ponadto niektóre z nich, np. HFC, podlegają szybkiemu rozkładowi w atmosferze, natomiast w przypadku PFC i SF_6 czas życia w atmosferze wynosi kolejno ok. 50 tysięcy lat i ok. 3 tysiące lat. Dodatkowo, występowanie fluorowanych gazów w atmosferze jest w pełni skutkiem ludzkiej działalności przemysłowej, ponieważ gazy te nie występują naturalnie w przyrodzie.

Należy zauważyć, że według raportu opublikowanego przez Joint Research Centre, Chiny oraz Stany Zjednoczone odpowiadają za największą emisję CO_2 w ujęciu globalnym. Jednocześnie kraje te produkują również najwięcej energii odnawialnej na świecie. Unia Europejska jako całość zajmuje trzecią lokatę [22].

W tabeli 1 przedstawiono zestawienie zmian zachodzących w emisji CO_2 (porównanie 2020 do 1990), gazów GHG (porównanie 2018 do 1990 oraz 2018 do 2005). Dane wskazują, iż w ciągu ok. 28-30 lat emisyjność gazów cieplarnianych oraz CO_2 wzrosła o około 57% we wszystkich sektorach. Największy wzrost emisyjności możemy zaobserwować w sektorze przemysłu energetycznego (w przypadku CO_2 to aż 71%, natomiast GHG 83%). Analizując natomiast rynek transportowy widoczny jest wzrost o 59% emisyjności CO_2 oraz 76% gazów GHG. Jednym z powodów jest niewątpliwie rosnąca liczba zarejestrowanych pojazdów samochodowych na świecie – w 2020 liczba ta przekroczyła 1 miliard (w porównaniu do 1990 zwiększyła się czterokrotnie).

Tabela 1. Zestawienie danych dotyczące emisji dwutlenku węgla i gazów cieplarnianych w ujęciu globalnym [60]

	2020 vs 1990 (CO ₂)	2018 vs 1990 (GHG)	2018 vs 2005 (GHG)
Przemysł energetyczny	+72%	+83%	+27%
Pozostałe spalanie przemysłowe	+58%	+62%	+26%
Gospodarstwa domowe	+1%	+2%	+5%
Transport	+59%	+76%	+26%
Pozostałe sektory	+97%	+48%	+21%
Wszystkie sektory	+58%	+57%	+23%

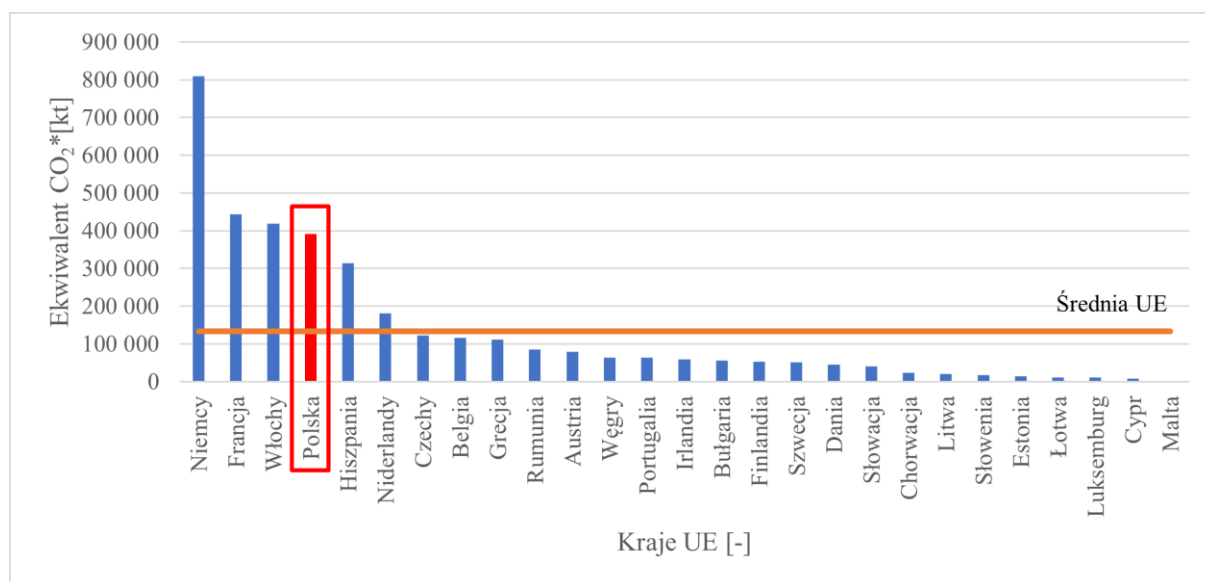
W tabeli 2 przedstawiono, jaki poziom miała emisja CO₂ i gazów cieplarnianych w analogicznych okresach w Polsce. Możemy zaobserwować, iż prawie we wszystkich kategoriach zmniejszyła się emisyjność CO₂, jak i GHG, zarówno w porównaniu do roku 1990, jak i do 2005. Wyjątkiem jest transport, gdzie emisyjność CO₂ w 2020 w porównaniu do roku 1990 wzrosła o 202%, natomiast GHG o 210%. Niewątpliwym powodem był wzrost liczby posiadanych pojazdów w Polsce, a także rosnąca mobilność społeczeństwa. W 1990 roku było zarejestrowanych 9 mln 41 tys. pojazdów samochodowych, natomiast w 2020 liczba ta wzrosła o 2,7 razy (do 25 mln 114 tys.). Wskaźnik zmotoryzowania społeczeństwa w Polsce w 1990 roku wynosił około 238 pojazdów samochodowych na 1000 mieszkańców, natomiast w 2020 roku aż 656 (wzrost o 276%). Jednak warto podkreślić, że w ujęciu globalnym wzrost ten nie był aż tak duży, ponieważ średnia emisyjność w sektorze transportu na świecie zwiększyła się o 68%. Patrząc natomiast na wskaźnik zmotoryzowania społeczeństwa na świecie w roku 1990 wynosił on 104 na 1000 mieszkańców, natomiast w 2020 – 177 (wzrost o 70%).

Tabela 2. Zestawienie danych dotyczące emisji CO₂ i gazów cieplarnianych w Polsce [60]

	2020 vs 1990 (CO ₂)	2018 vs 1990 (GHG)	2018 vs 2005 (GHG)
Przemysł energetyczny	-45%	-32%	-10%
Pozostałe spalanie przemysłowe	-25%	-20%	-4%
Gospodarstwa domowe	0%	+4%	-1%
Transport	+202%	+210%	+83%
Pozostałe sektory	-11%	-30%	-3%
Wszystkie sektory	-21%	-17%	+1%

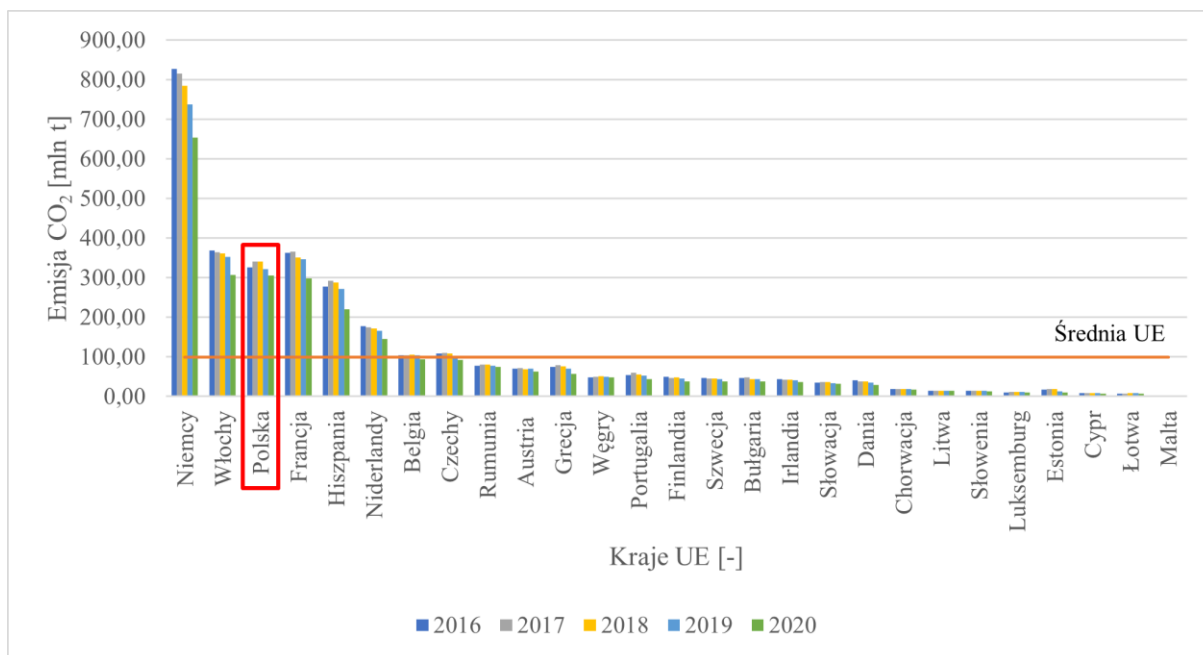
Miarą używaną do wyrażenia potencjału globalnego ocieplenia (GWP) jest ekwiwalent dwutlenku węgla (CO₂e). Umożliwia on porównanie emisji różnych gazów cieplarnianych we wspólnej skali, w jeden ustandaryzowany sposób, pod kątem ich zdolności do zatrzymywania

ciepła i pozostawania w atmosferze przez określony czas. Na podstawie danych dotyczących emitowania ekwiwalentu CO₂ w 2019 roku, Polska znajduje się w czołówce i zajmuje 4 miejsce na tle Unii Europejskiej, emitując ok. 10,81% ekwiwalentu CO₂ (rys. 4). Pierwsze miejsce zajmują Niemcy, osiągając sześciokrotnie większą emisję ekwiwalentu CO₂ niż średnia w Unii Europejskiej (133 894,19 kt CO₂e) [85]. Drugie miejsce przypada Francji - 12,25% emisji ekwiwalentu Europejskiego, a na trzecim miejscu natomiast znajdują się Włochy, osiągając ok. 11,57% całkowitej emisji gazów cieplarnianych w państwach Unii Europejskiej.



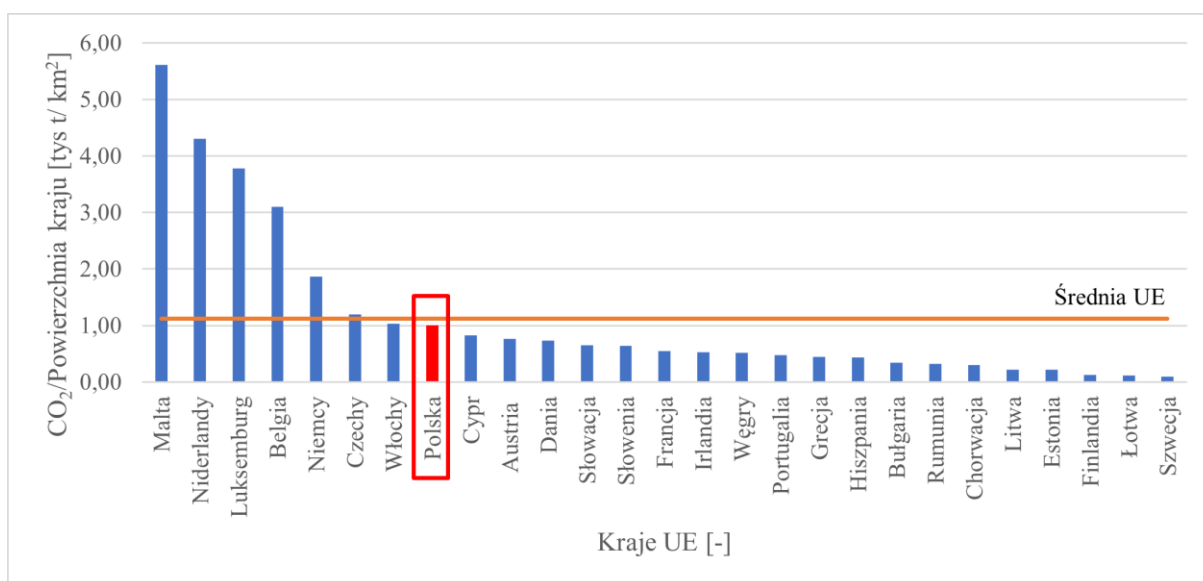
Rys. 4. Emisja gazów cieplarnianych w UE w 2019 roku (gdzie * oznacza CO₂ oraz N₂O, CH₄, HFC, PFC, SF₆, NF₃ wyrażone w ekwiwalencie CO₂) [85]

Według sprawozdania z 2021 roku Międzyrządowego Zespołu ds. Zmian Klimatu (organizacji mającej na celu dostarczenie naukowej informacji na temat zmiany klimatu i jego wpływu na gospodarkę i społeczeństwo) emisja gazów cieplarnianych z działalności człowieka odpowiada za wzrost temperatury o ok. 1,1°C od początku XX wieku. Na rysunku 5 przedstawiono emisyjność w latach 2016–2020 w poszczególnych krajach Unii Europejskiej. Możemy zaobserwować, że państwem który produkuje najwięcej CO₂ w Unii Europejskiej są Niemcy, Polska natomiast zajmuje 3 miejsce, produkując trzykrotnie więcej CO₂ niż średnia UE z 10 lat (119,3 mln ton). Przez zastosowanie działań, mających na celu ograniczenie emisyjności CO₂, wprowadzonych przez Unię Europejską, możemy zaobserwować tendencję spadkową w poszczególnych krajach. Jednakże wartości w 2019 oraz 2020 mogą nie być wiarygodne z powodu, iż w tym okresie zostało wprowadzone ograniczenie mobilności, wynikające z wybuchu pandemii COVID-19.



Rys. 5. Emisja CO₂ w krajach Unii Europejskiej w latach 2016-2020 [126]

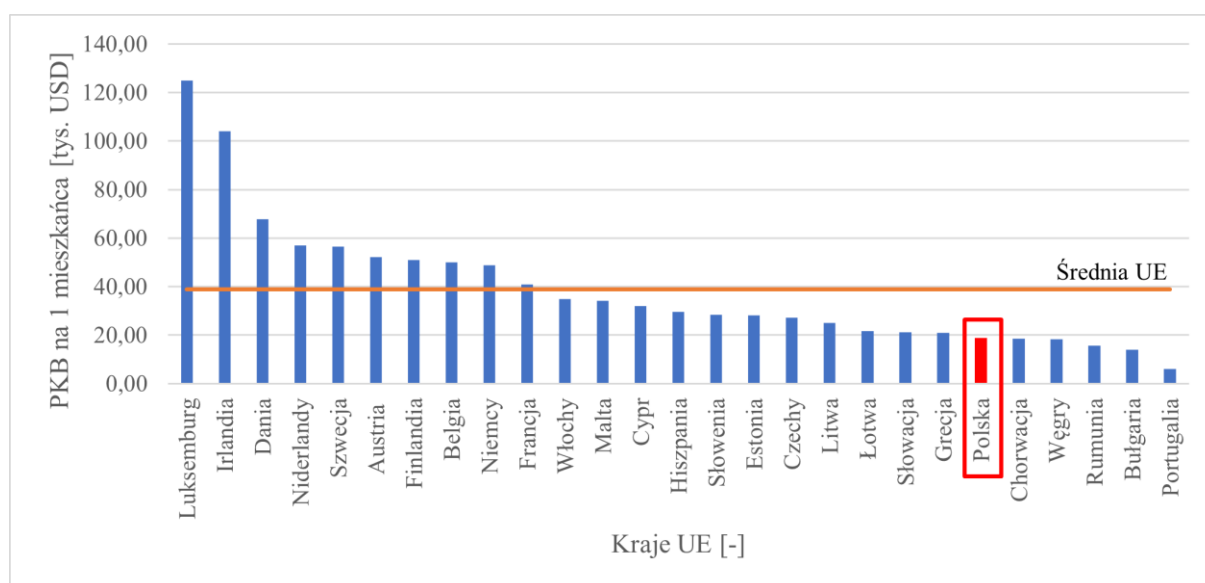
Biorąc pod uwagę powierzchnię kraju jak i wielkość danego państwa, ranking państw pod względem emisyjności CO₂ do atmosfery całkowicie się zmienia: Polska zajmuje 8 miejsce, natomiast liderem jest Malta, emitująca na 1 km² 5,62 tys. ton CO₂ (przekraczając 5 krotnie średnią UE 1,12) (rys. 6).



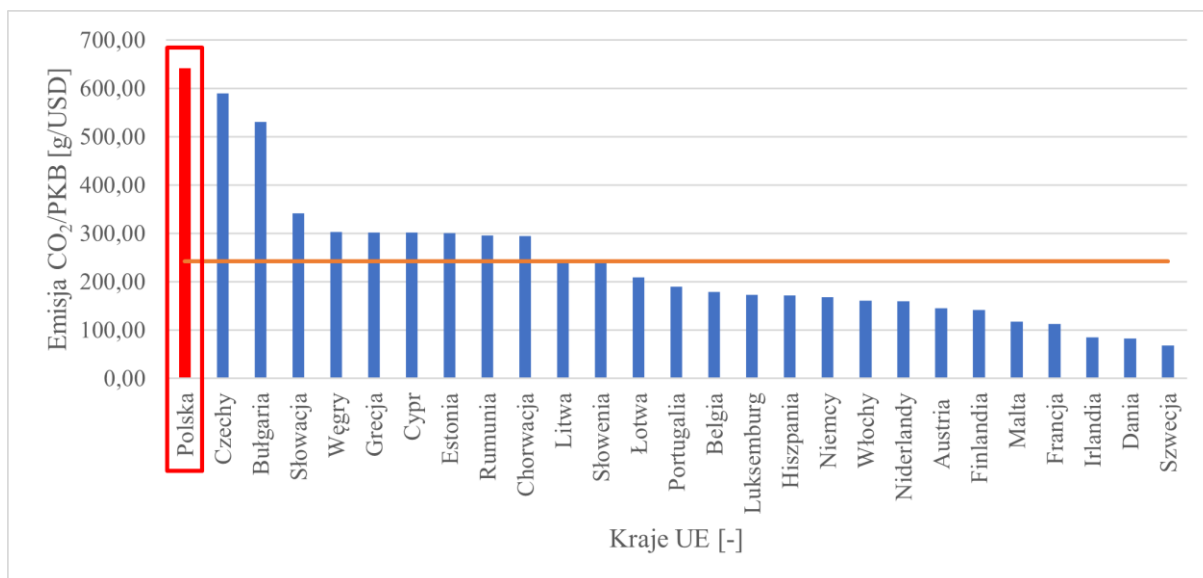
Rys. 6. Względna emisja CO₂ w poszczególnych krajach Unii Europejskiej w zależności od powierzchni państwa w roku 2020 [126]

Inną miarą porównania emisyjności w poszczególnych krajach jest wyznaczenie wskaźnika emisyjności CO₂ w odniesieniu do Produktu Krajowego Brutto (PKB), czyli określenie, jak aktywność gospodarcza wpływa na środowisko. Kraje rozwinięte mają niższą intensywność emisji CO₂ na jednostkę PKB w porównaniu z krajami rozwijającymi się. Wynika to z większej efektywności energetycznej, zaawansowanych nowoczesnych

technologii oraz większego udziału sektorów usługowych, które są mniej emisyjne, niż intensywna produkcja przemysłowa. Kraje, które aktywnie inwestują w odnawialne źródła energii i prowadzą politykę ograniczającą emisję CO₂, zwykle osiągają lepsze wyniki w zakresie emisyjności CO₂ na jednostkę PKB. Według danych World Development Indicators z 2022, Polska ma dziewiąte miejsce w rankingu PKB w odniesieniu do pozostałych państw UE [125]. Natomiast w przeliczeniu PKB na 1 mieszkańca Polska zajmuje 22 miejsce w UE, osiągając ok. 119 tys. USD na osobę. Kraje z najwyższym PKB na osobę to: Luksemburg i Irlandia, osiągając ponad 100 tysięcy USD na osobę (rys. 7). Jednakże biorąc pod uwagę ww. wskaźnik, na 1 USD Polska produkuje około 640 g CO₂ do atmosfery (1 lokata w krajach UE), na drugim miejscu jest Bułgaria. Najlepszą efektywność gospodarczą ma natomiast Szwecja, gdzie na 1 USD produkuje się jedynie 68 g CO₂ (tj. 7 razy mniej niż Polska) (rys. 8). Taki wynik pokazuje, że Polska ma najniższą efektywność w porównaniu do pozostałych krajów Unii Europejskiej. Wpływa na to między innymi trwające uzależnienie od węgla kopalnianego i brak rozwiązań, które przyczyniłyby się do większego korzystania z odnawialnych źródeł energii.

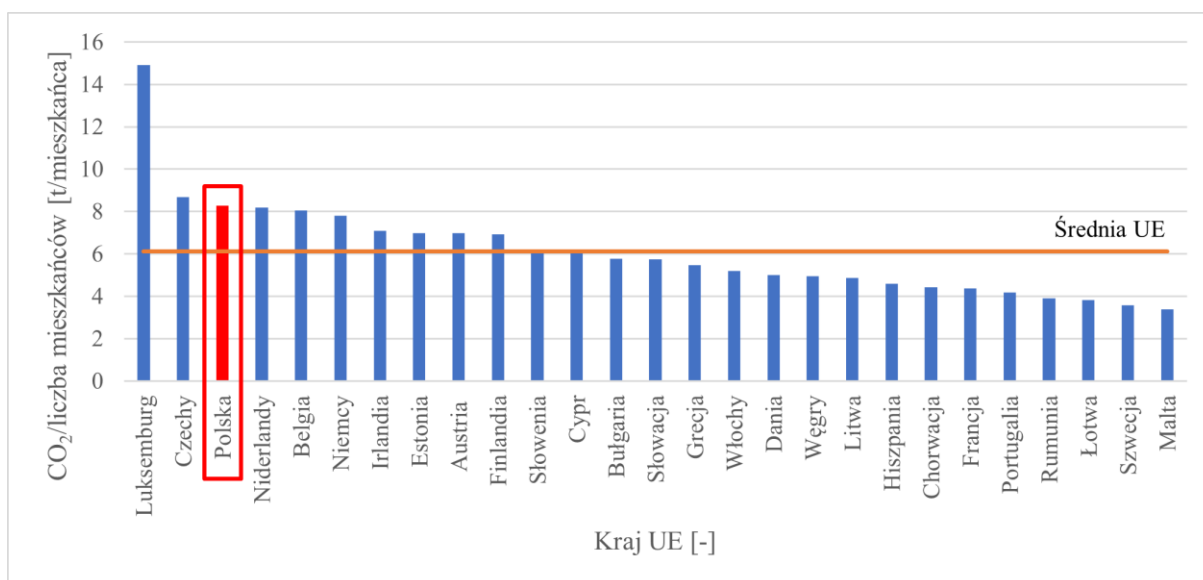


Rys. 7. Produkt Krajowy Brutto w przeliczeniu na 1 mieszkańca w 2022 roku w krajach Unii Europejskiej [126]



Rys. 8. Wpływ efektywności gospodarczej krajów Unii Europejskiej na środowisko w 2020 roku [126]

Biorąc pod uwagę kolejny aspekt, a mianowicie emisję CO₂ w przeliczeniu na 1 mieszkańca, Polska również znajduje się wysoko w rankingu, zajmując 3 miejsce, tj. produkując około 8 ton rocznie CO₂ w przeliczeniu na 1 mieszkańca (średnia UE natomiast to 6,12 ton). Pierwszą lokatę zajmuje Luksemburg, emitując około 15 ton rocznie (2,5 razy więcej niż UE) – rysunek 9. Wysokie wartości tego wskaźnika wykazują bardziej emisyjny styl życia oraz wzorce konsumpcji w danym kraju. Kraje bogatsze mają zwykle wyższe emisje na osobę, ze względu na większe zużycie i konsumpcję energii. Porównując kraje UE, Polska posiada porównywalną wartość emisji CO₂ w przeliczeniu na 1 mieszkańca w stosunku do krajów takich jak Belgia, Niderlandy, czy Niemcy, które mają dużo wyższe PKB w przeliczeniu na jednego mieszkańca.

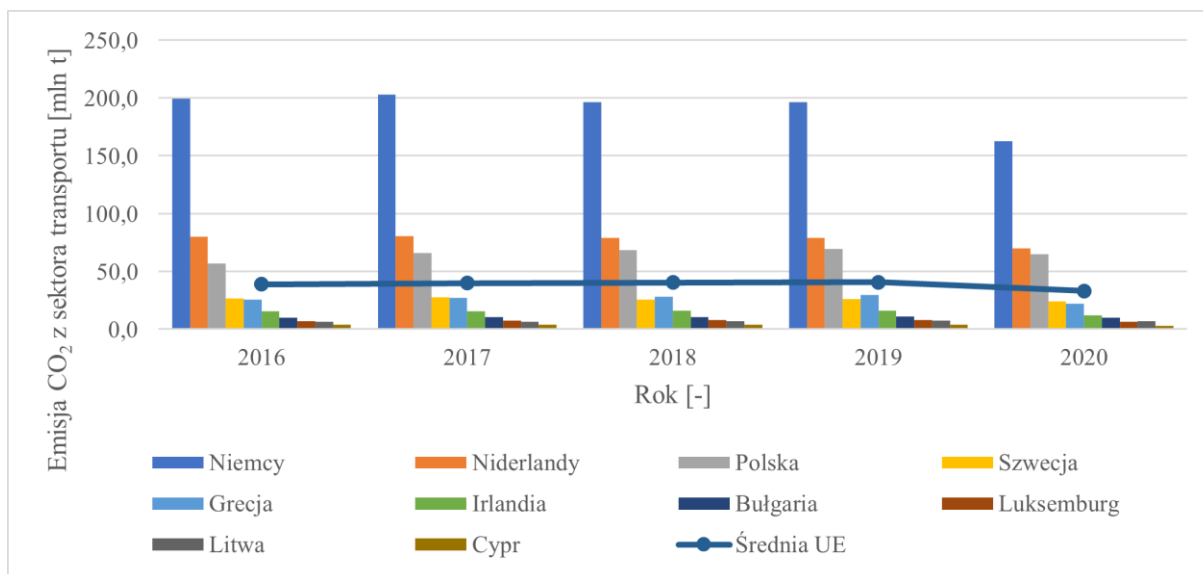


Rys. 9. Emisja CO₂ w przeliczeniu na 1 mieszkańca w 2020 roku [126]

Emisje CO₂ według sektorów: UE-27 [mtn t]

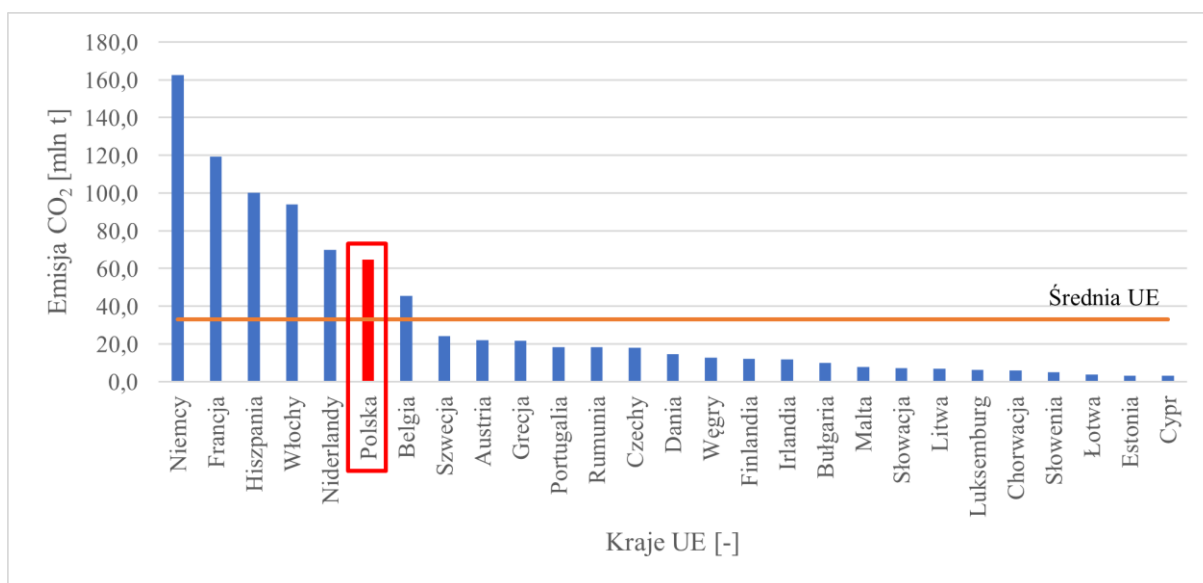
Lata [-]	Przemysł energetyczny	Produkcja i Budownictwo	Transport	Pozostałe sektory
2010	1240,0	480,0	910,0	640,0
2011	1230,0	460,0	900,0	570,0
2012	1210,0	440,0	870,0	570,0
2013	1160,0	430,0	860,0	570,0
2014	1090,0	410,0	870,0	500,0
2015	1100,0	420,0	890,0	520,0
2016	1070,0	420,0	910,0	530,0
2017	1060,0	430,0	940,0	520,0
2018	1010,0	430,0	940,0	510,0
2019	890,0	420,0	950,0	500,0
2020	780,0	400,0	770,0	500,0

Porównując dane z 27 krajów UE dotyczące emisji CO₂ pochodzącej z sektora transportu można zauważyć, że średnia emisja CO₂ krajów Unii Europejskiej wynosi 38,12 miliona ton. Aby wykazać poziom istniejących różnic, na rysunku 11 zestawiono jedynie wybrane kraje. Wybrano te, które zajmowały wysokie miejsca w rankingu emisyjności CO₂ (tj. Niemcy, Niderlandy, Polska, Luksemburg Irlandia, Grecja, Bułgaria). Porównano je z krajami o najniższej emisyjności tj. Litwa, Cypr, Szwecja. Emisja w wybranych państwach, na przestrzeni 10 lat (2010-2020) wyniosła odpowiednio: Niemcy – średnio 189,8 mln ton, Niderlandy tj. 77,6 mln ton. Polska natomiast osiągnęła wynik 65 mln ton emisji CO₂ do atmosfery. Według danych z Dyrekcji Generalnej ds. Mobilności i Transportu, działającej przy Komisji Europejskiej, w latach 2010-2015 Polska miała o 30% wyższą emisyjność niż średnia UE, jednak od 2016 roku odnotowuje tendencję wzrostową (2016 wzrost o 46%, 2020 wzrost o 97%). Najniższą emisyjność z sektora transportu w całej Unii Europejskiej w latach 2016-2020 miał Cypr - 3,1 mln ton (średnio 3,4 mln ton) [59].



Rys. 11. Emisja CO₂ z transportu z uwzględnieniem zbiorników międzynarodowych [59]

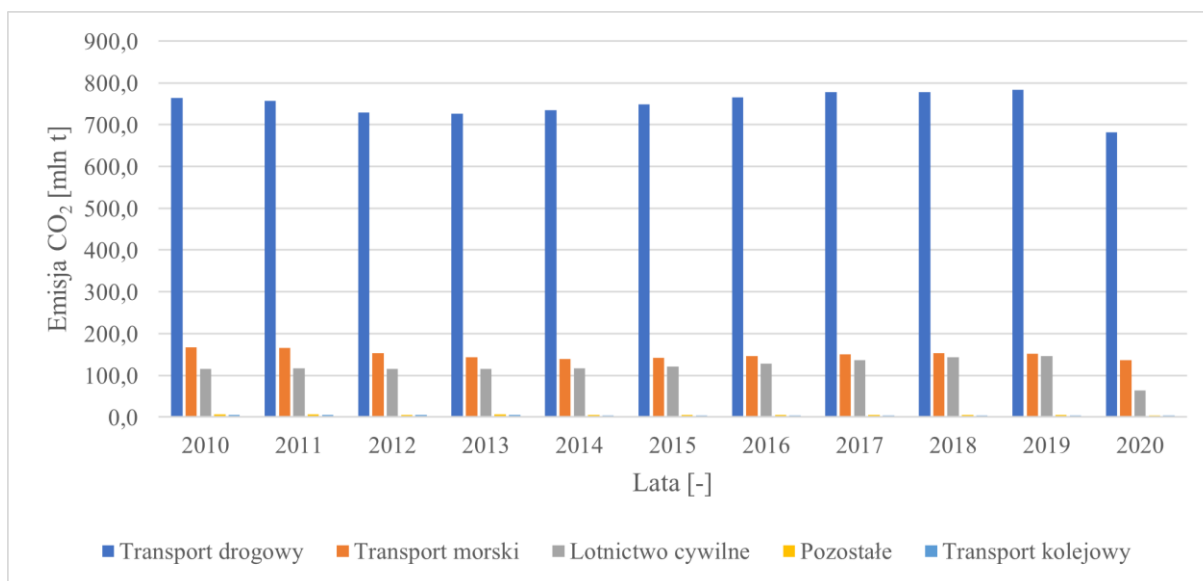
Analizując państwa zaprezentowane na rysunku 12, w 2020 roku wysoko w zestawieniu znalazły się również takie kraje jak Francja – 119,4 mln ton oraz Hiszpania – 100 mln ton. Najniższą emisyjność miały kraje takie jak: Łotwa (3,9 mln ton), Estonia (3,2 mln ton) oraz Cypr (3,1 mln ton). Biorąc pod uwagę jak wyglądała zmiana emisji CO₂ w stosunku do roku poprzedniego (2019), to średnio w Europie wszystkie kraje UE zmniejszyły swoją emisyjność o ok. 20%, Natomiast patrząc sumarycznie, kraje Unii Europejskiej wyemitowały o 23% mniej CO₂ do atmosfery z sektora transportu, tj. 203,5 mln ton.



Rys. 12. Emisja CO₂ z transportu w krajach Unii Europejskiej w 2020 roku [59]

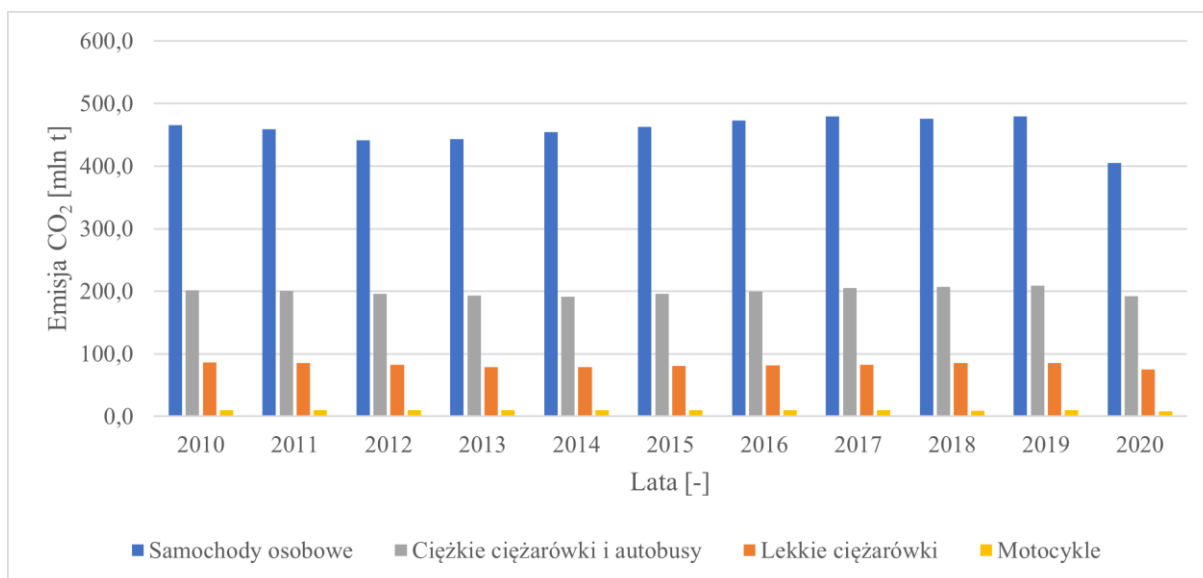
Według danych podanych przez Europejską Agencję Środowiskową, w 2020 roku około 76% emisji CO₂ w UE z sektora transportu stanowił transport drogowy (jedna piąta całkowitej emisji CO₂ w Unii Europejskiej) – rysunek 13. Drugie miejsce zajął transport morski (15%). Zaobserwować można spadek wartości w roku 2020 o 13% w porównaniu do roku

2019, do czego mogły się przyczynić wprowadzone ograniczenia w przemieszczaniu się w trakcie pandemii COVID-19.



Rys. 13. Emisja CO₂ w zależności od rodzaju transportu w 2020 roku [126]

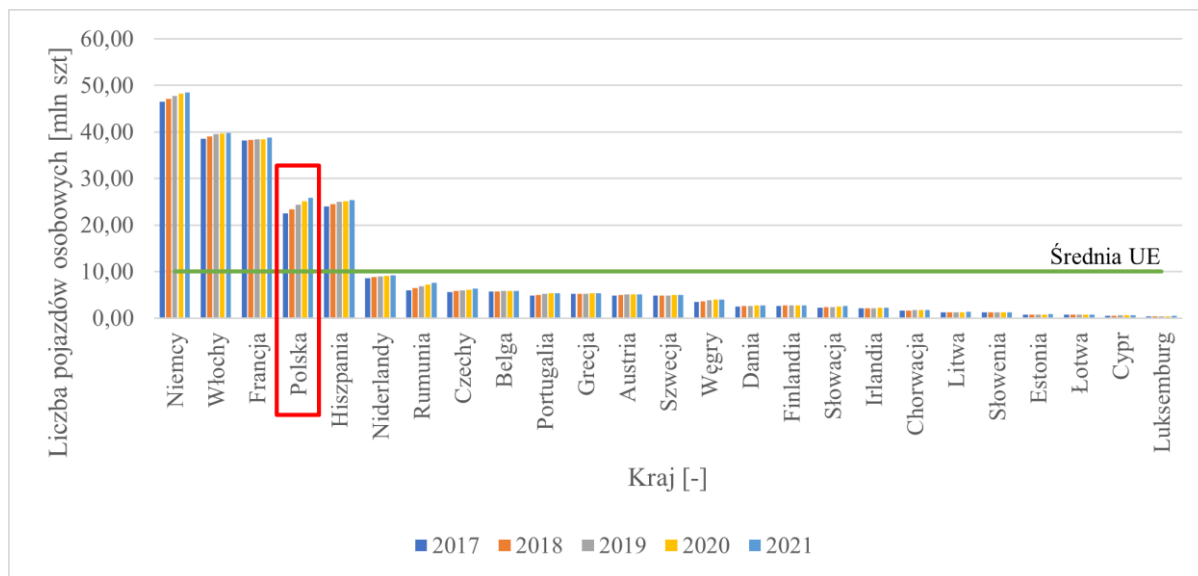
Na rysunku 14 przedstawiono dane dotyczące podziału emisyjności pochodzącej z transportu drogowego, w zależności od rodzaju pojazdów. Źródłem 60% całkowitej emisji CO₂ są samochody osobowe, 28,2 % pojazdy ciężarowe, 11% lekkie samochody dostawcze i 1,2% motocykle [84]. Do głównych gazów cieplarnianych, emitowanych przez pojazdy silnikowe można zaliczyć dwutlenek węgla (CO₂), metan (CH₄) oraz podtlenek azotu (N₂O) [12], [99].



Rys. 14. Emisje CO₂ z transportu drogowego, według środka transportu: UE-27 w 2020 roku [59]

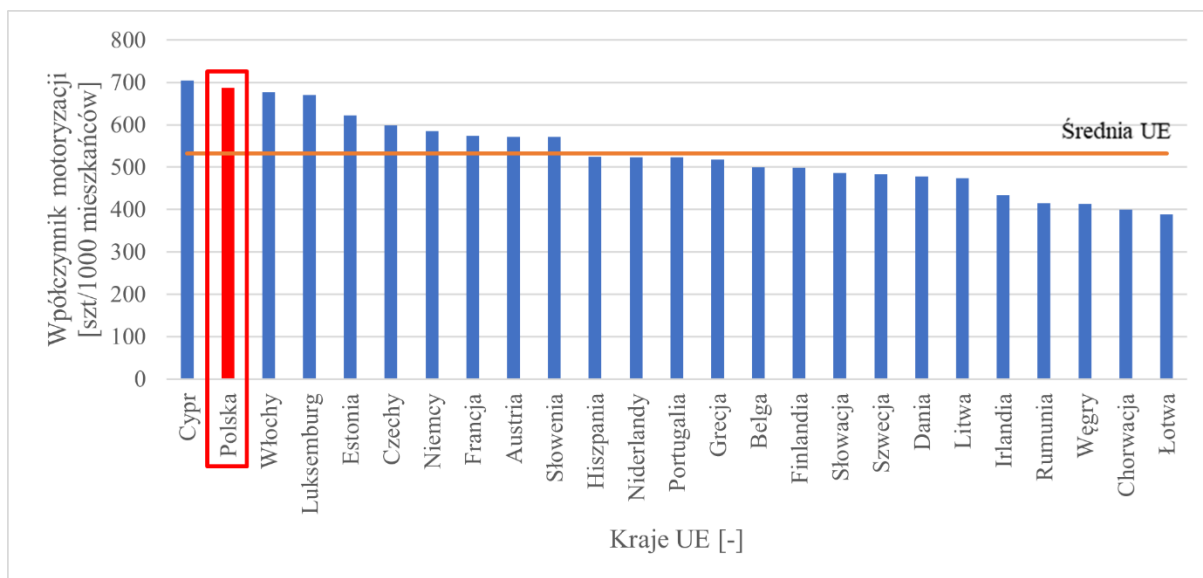
Analizie poddano również liczbę zarejestrowanych pojazdów samochodowych. W 2021 roku w całej Unii Europejskiej było ich około 250 mln, z czego 19% w Niemczech (48,54 mln).

Polska zajęła 4 miejsce w rankingu, z sumą 25,87 mln pojazdów samochodowych (10%) (rys. 15). Średnia liczba pojazdów w Europie to około 10 mln. Można zauważyć, że każdego roku przybywa około 3,7 mln pojazdów w całej Unii Europejskiej. W Polsce wzrost ten wynosi około 850 tysięcy.



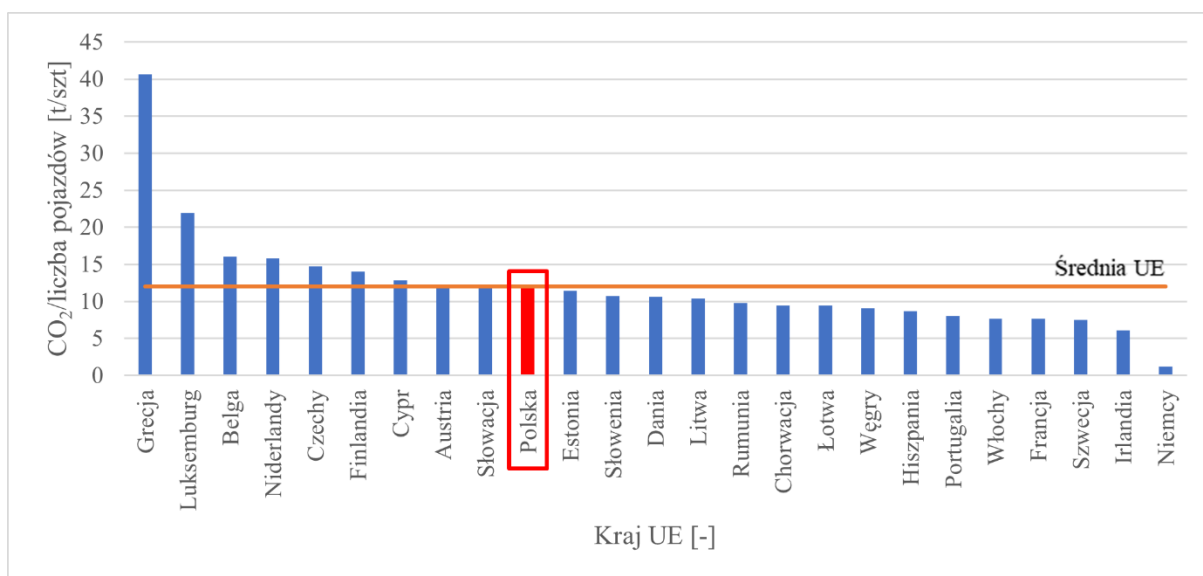
Rys. 15. Liczba samochodów osobowych w UE w latach 2017-2021 [91]

Miarą, która określa poziom motoryzacji oraz poziom rozwoju transportu i gospodarki w danym kraju jest także współczynnik motoryzacji. Na rysunku 16 przedstawiono wartość tego współczynnika w poszczególnych krajach Unii Europejskiej w 2021 roku. Na pierwszym miejscu w rankingu znalazł się Cypr, osiągając 705 pojazdów osobowych na 1000 mieszkańców, natomiast na drugim miejscu znalazła się Polska, z wynikiem 686 pojazdów, oznaczającym wartość większą o 29% niż średnia krajów Unii Europejskiej, wynosząca 533 pojazdów. Na trzecim miejscu znalazły się Włochy (677 pojazdów), a na czwartym Luksemburg (671 pojazdów). Najniższy, współczynnik osiągnęła natomiast Łotwa (389 pojazdów). Niemcy, pomimo największej liczby pojazdów w Europie, zajęły 7 miejsce (585 pojazdów).



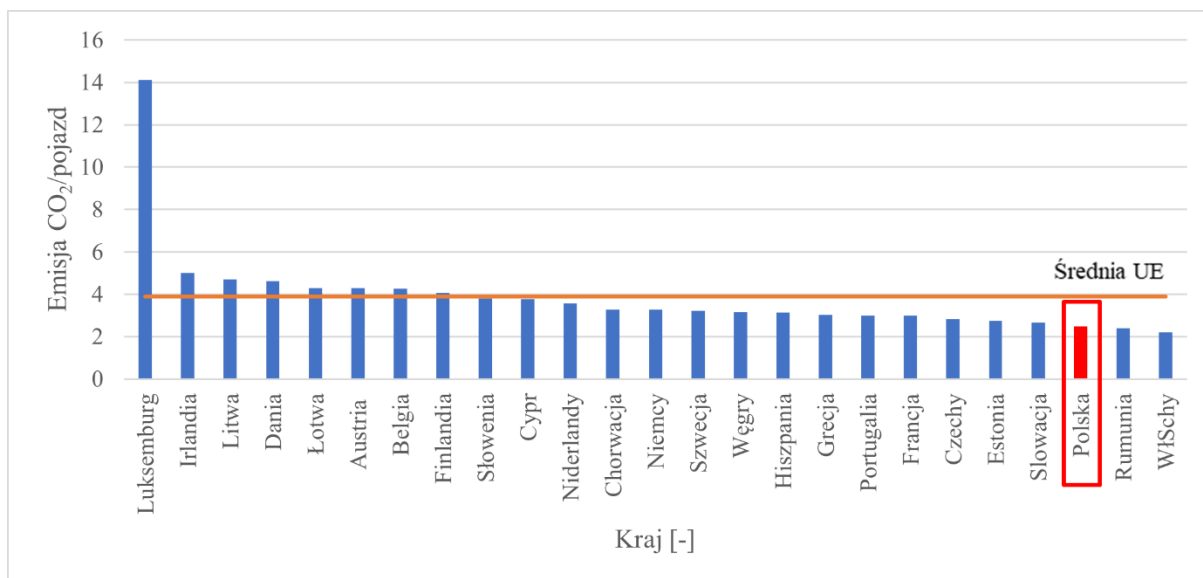
Rys. 16. Współczynnik motoryzacji w krajach Unii Europejskiej w 2021 roku [126]

Ważnym aspektem jest również informacja o tym, jak liczba pojazdów przekłada się na emisyjność CO₂ w danym kraju. W tym przypadku Grecja, pomimo tego, iż posiada jedynie 5,41 mln pojazdów samochodowych (zajmując 11 miejsce w rankingu dotyczącym liczby posiadanych pojazdów) oraz dla której współczynnik motoryzacji wyniósł 518 pojazdów na 1000 mieszkańców (14 miejsce w rankingu), ma najwyższy poziom emisji CO₂ w przeliczeniu na 1 pojazd – 40,62 ton rocznie (rys. 17). Natomiast Niemcy, pomimo największej liczby pojazdów w UE, mają najniższą emisyjność, wynoszącą jedynie 1,17 t (34 razy mniej niż Grecja). Drugie miejsce w zestawieniu zajmuje Luksemburg (21,96 tony), natomiast Polska zajmuje 10 miejsce – 11,78 tony.



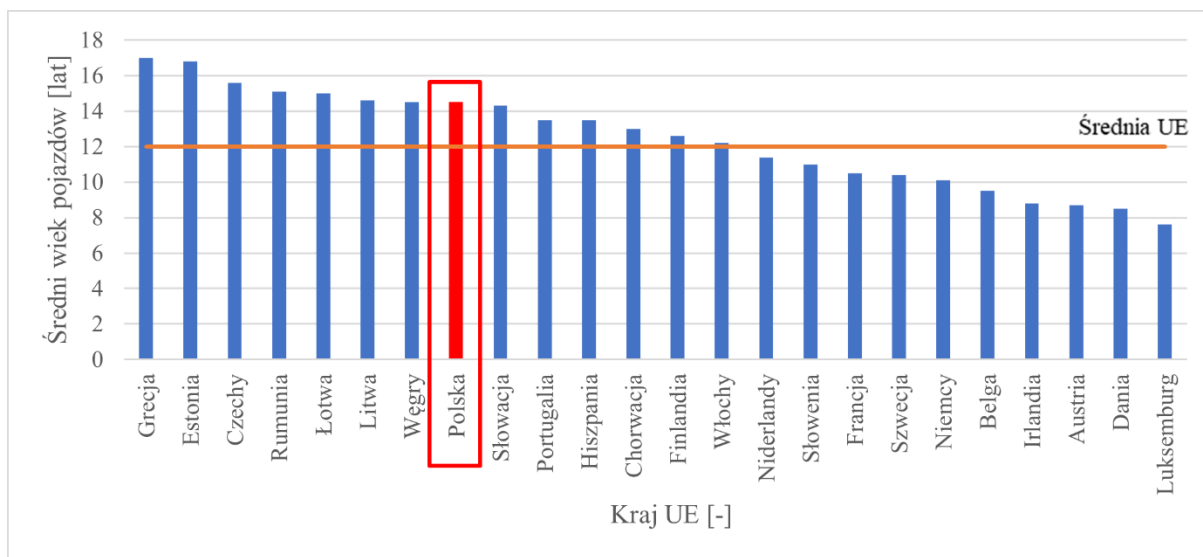
Rys. 17. Całkowita emisja CO₂ w zależności do liczby pojazdów osobowych w krajach Unii Europejskiej w roku 2020 [126]

Oceniając natomiast dane dotyczące emisji pochodzącej z sektora transportu – najwyższą emisyjność w 2020 roku, w przeliczeniu na 1 pojazd samochodowy miał Luksemburg (14,11 t), na drugim miejscu znalazła się Irlandia (5 t) i a na trzecim Litwa (4,7 t). Polska natomiast zajęła 23 pozycję w rankingu z wynikiem 2,46 tony na pojazd. Średnia wartość dla krajów Unii Europejskiej wynosiła 3,88 tony (rys. 18).



Rys. 18. Emisja CO₂ z sektora transport w zależności do liczby pojazdów osobowych w krajach Unii Europejskiej w 2020 roku [126]

Ważnym czynnikiem jest również średni wiek pojazdów. Na rysunku 19 przedstawiono wyniki dla poszczególnych krajów UE w 2021 roku. Najstarsze pojazdy posiadają Grecy (17 lat), natomiast najnowsze samochody posiadają mieszkańcy Luksemburga (7,6 roku). Średnia Unii Europejskiej to 12 lat, natomiast Polska zajmuje 8 pozycję w rankingu – średni wiek pojazdów samochodowych wynosi 14,5 roku. Patrząc na kraje, które osiągnęły wysoką pozycję w rankingu można zauważyć, że średni wiek pojazdów w Irlandii to 8,8 roku, natomiast na Litwie 14,6 roku.



Rys. 19. Średni wiek pojazdu w UE w 2021 roku [126]

Kolejnym kluczowym aspektem jest odsetek pojazdów starszych niż 10 lat w krajach Unii Europejskiej. Według danych z 2021 roku krajem, który ma najwięcej pojazdów starszych niż 10 lat, tj. 87% jest Rumunia (6,5 miliona). Na drugim miejscu jest Grecja z wynikiem (84% - 4,5 miliona). Polska znajduje się na trzecim miejscu, gdzie 81% pojazdów (około 21 milionów) to pojazdy starsze niż 10 lat. Sprzyja temu fakt, że wiele z poruszających się po polskich drogach pojazdów jest sprowadzana z krajów Unii Europejskiej a jedynie 1,3 mln wśród nich to pojazdy do 3 lat [126]. Luksemburg i Irlandia mają najniższą liczbę pojazdów w tej kategorii wiekowej odpowiednio 25% i 37%.

Istotnym czynnikiem jest także źródło zasilania pojazdu. Według raportu ACEA „Pojazdy w użyciu” z 2023 r. [92] około 51,1% wszystkich samochodów w Unii Europejskiej jest napędzane benzyną. Olej napędowy jako paliwo dotyczy ok. 41,9% pojazdów, gaz LPG – 2,5%. Pojazdów typu HEV (z ang. Hybrid Electric Vehicle) jest 2,3%, BEV (z ang. Battery Electric Vehicle) jedyne 0,8%, natomiast 0,7% to PHEV (z ang. Plug-in Hybrid Electric Vehicle) - szczegóły prezentuje tabela 3. W Luksemburgu, będącym wysoko w rankingach dotyczących emisyjności CO₂, podział pojazdów według kryterium źródła zasilania jest następujący: benzyna (43,4%), olej napędowy (49%), BEV (1,9). W Irlandii: benzyna (36,8%), olej napędowy (57,2%), BEV (0,9), PHEV (0,8%), oraz HEV (4,0%). Na Litwie: benzyna (25,2%), olej napędowy (67,8%), HEV (3%). Natomiast Grecja (z uwagi na ranking dotyczący wieku pojazdów) ma aż 90,1% pojazdów zasilanych benzyną i jedynie 8,6% olej napędowy. Patrząc natomiast na Polskę, samochody zasilane benzyną stanowią 44,7%, olejem napędowym 39,9%, LPG – 13,4%, a HEV to jedyne 1,7%. Zauważyć należy, że najmniejszy jest udział pojazdów BEV i PHEV, wynoszący po 0,1% na każdy typ pojazdu [91].

Na podstawie przedstawionych statystyk zaobserwowano, że kraje Unii Europejskiej nadal w głównej mierze uzależnione są od pojazdów zasilanych paliwami kopalnianymi. W 2021 r. w Unii Europejskiej zarejestrowano około 1,74 mln nowych samochodów elektrycznych (pojazdów BEV i PHEV), które stanowiły 18% wszystkich rejestracji nowych samochodów. W 2022 r. liczba ta nadal rosła, przy czym samochody elektryczne stanowiły około 21,6% (prawie 2 miliony) rejestracji nowych samochodów w UE [35].

Tabela 3. Podział pojazdów według źródła zasilania w krajach UE w 2021 roku [91]

Kraj	Benzyna	ON	BEV	PHEV	HEV	CNG	LPG	Inne
Austria	42,8%	52,9%	1,5%	0,0%	2,7%	0,1%	0,1%	0,0%
Belgia	51,7%	42,8%	0,9%	2,0%	1,7%	0,3%	0,3%	0,3%
Chorwacja	40,0%	55,5%	0,2%	0,1%	0,8%	0,0%	3,4%	0,0%
Cypr	75,5%	21,9%	0,1%	0,0%	2,4%	0,0%	0,0%	0,1%
Czechy	65,4%	32,3%	0,1%	0,1%	0,8%	0,3%	0,2%	0,7%
Dania	64,9%	28,6%	2,4%	2,8%	1,3%	0,0%	0,0%	0,0%
Estonia	56,0%	40,9%	0,3%	0,0%	2,4%	0,3%	0,0%	0,0%
Finlandia	68,4%	27,3%	0,8%	2,8%	0,0%	0,5%	0,0%	0,2%
Francja	40,1%	55,2%	1,0%	0,8%	2,5%	0,5%	0,0%	0,0%
Niemcy	63,9%	30,5%	1,3%	1,2%	2,3%	0,2%	0,7%	0,0%
Grecja	90,1%	8,6%	0,0%	0,0%	0,6%	0,1%	0,0%	0,5%
Węgry	64,2%	31,7%	0,5%	0,5%	2,4%	0,1%	0,7%	0,0%
Irlandia	36,8%	57,2%	0,9%	0,8%	4,0%	0,0%	0,0%	0,3%
Włochy	44,7%	42,9%	0,3%	0,0%	2,6%	2,5%	7,0%	0,0%
Łotwa	32,9%	65,0%	0,3%	0,0%	0,7%	0,0%	0,1%	0,9%
Litwa	25,2%	67,8%	0,3%	0,1%	3,0%	0,0%	0,2%	3,4%
Luksemburg	43,4%	49,0%	1,9%	1,9%	3,7%	0,0%	0,0%	0,1%
Niderlandy	79,5%	11,1%	2,8%	1,5%	3,9%	0,1%	1,1%	0,0%
Polska	44,7%	39,9%	0,1%	0,1%	1,7%	0,0%	13,4%	0,1%
Portugalia	36,8%	59,1%	0,8%	0,9%	1,5%	0,0%	0,0%	1,0%
Rumunia	48,7%	49,5%	0,2%	0,0%	1,1%	0,0%	0,0%	0,5%
Słowacja	50,5%	44,0%	0,1%	0,1%	1,3%	0,1%	1,9%	1,9%
Słowenia	46,5%	51,1%	0,4%	0,1%	1,2%	0,0%	0,5%	0,1%
Hiszpania	39,5%	56,9%	0,3%	0,3%	2,7%	0,1%	0,3%	0,0%
Szwecja	51,8%	34,6%	2,2%	3,8%	3,1%	0,8%	0,0%	3,7%
UE	51,1%	41,9%	0,8%	0,7%	2,3%	0,6%	2,5%	0,2%

Na podstawie przeprowadzonych analiz dotyczących emisyjności, w tabeli 4 przedstawiono zestawienie zbiorcze dla wybranych państw Unii Europejskiej. Wybrano państwa, które w trakcie wcześniejszych badań uzyskały wysokie miejsca w rankingach. Niemcy produkują do atmosfery najwięcej CO₂ w całej Unii Europejskiej (653 mln ton), tym samym zajmują również pierwsze miejsce w zakresie jego emisji z sektora transportu (162,5 mln ton rocznie) oraz w zakresie liczby posiadanych pojazdów osobowych (48,54 mln), które stanowią 1/5 wszystkich pojazdów osobowych w UE. Patrząc natomiast na pozostałe statystyki,

prezentujące np. jak PKB, liczba ludności czy powierzchnia kraju przekłada się na emisyjność CO₂, ranking całkowicie się zmienia i w żadnym z nich kraj ten nie zajmuje wysokich pozycji. Biorąc przykładowo pod uwagę współczynnik motoryzacji, Niemcy zajmują 7 miejsce, a 19 miejsce w odniesieniu do wieku pojazdów – średnia to 10,1 roku. Jedynie 43,3% niemieckich pojazdów to pojazdy starsze niż 10 lat.

Luksemburg jest krajem, który ma najwyższe PKB w przeliczeniu na 1 mieszkańca (125,01 USD). Całkowita emisja CO₂ jest 67 razy mniejsza niż w Niemczech, natomiast ta, która pochodzi z sektora transportu – 25 razy. Liczba pojazdów osobowych jest 110 razy mniejsza niż w Niemczech i pomimo tego, iż posiadają najmłodszą flotę pojazdów w Unii Europejskiej (średnio 7,6 lat) oraz faktu, iż jedynie 25,1% pojazdów jest starszych niż 10 lat, to mają najwyższą emisyjność CO₂ z sektora transport w przeliczeniu na 1 pojazd, wynoszącą – 14,11 t. Współczynnik motoryzacji w Luksemburgu wynosi 671, lokując go na 4 miejscu.

Grecja ma 12-krotnie mniejszą emisyjność niż Niemcy, a w odniesieniu do PKB na mieszkańca jest na 21 miejscu wśród krajów Unii Europejskiej. Zauważyć należy, że jest bezwzględnym liderem w zakresie posiadania najstarszych pojazdów, gdzie ich średni wiek wynosi 17 lat, a procent pojazdów starszych niż 10 lat wynosi 83,8 % (2 miejsce w Unii Europejskiej). Współczynnik motoryzacji tego kraju wynosi 518. Polska również znajduje się wysoko w rankingach, zajmując 3 miejsce w odniesieniu do emisji CO₂ i ma jedynie 1,9 razy mniejszą emisyjność niż Niemcy. Emisja z sektora transportu jest na 6 miejscu w Unii Europejskiej. Patrząc natomiast na PKB, jest niżej niż Grecja, zajmując 22 miejsce, przez co posiada wysoką wartość współczynnika emisji CO₂ do PKB. Względem liczby posiadanych pojazdów jest 4 w UE, a współczynnika motoryzacji – 2. Średni wiek pojazdów poruszających się na polskich drogach to 14,5 lat. Pojazdy powyżej 10 lat stanowią 80,6% (3 miejsce w Unii Europejskiej).

Tabela 4. Zestawienie danych dotyczących przedstawionych analiz dla wybranych krajów UE

	Niemcy	Luksemburg	Irlandia	Litwa	Grecja	Polska	Średnia UE
Produkt Krajowy Brutto na 1 mieszkańca [tys. USD/1 mieszkańca]	48,72	125,01	103,98	25,06	20,87	18,69	38,81
Całkowita emisja CO ₂ [mln t]	653,00	11,3	36,33	13,82	56,93	304,87	99,56
Emisja CO ₂ z sektora transport [mln t]	162,5	6,3	11,8	6,8	21,8	64,8	32,9
Emisja CO ₂ na 1 mieszkańca [t/1 mieszkańca]	7,79	14,9	7,09	4,88	5,46	8,27	6,12

	Niemcy	Luksemburg	Irlandia	Litwa	Grecja	Polska	Średnia UE
Emisja CO ₂ na powierzchnie kraju [tys. t/km ²]	1,87	3,78	0,53	0,22	0,44	1,00	1,12
Emisja CO ₂ na PKB [kg/USD]	167,97	173,14	84,76	242,53	302,70	640,94	242,51
Liczba pojazdów [mln szt.]	48,54	0,44	2,27	1,33	5,41	25,87	9,98
Współczynnik motoryzacji [szt./1000 mieszkańców]	585	671	434	473	518	686	533
Emisja CO ₂ na 1 pojazd [t/pojazd]	1,17	21,96	6,10	10,42	40,64	11,78	11,99
Emisja CO ₂ z sektora transport na 1 pojazd [t/pojazd]	3,27	14,11	5,01	4,70	3,04	2,47	3,88
Wiek pojazdów [lata]	10,1	7,6	8,8	14,6	17,0	14,5	12,0
Procent pojazdów starszych niż 10 lat [%]	43,3	25,1	37,1	77,5	83,8	80,6	59,0

Przedstawiona analiza ukazuje znaczne zróżnicowanie pomiędzy poszczególnymi państwami. Z jednej strony wynika ono z różnego poziomu rozwoju gospodarczego, wykorzystywanych źródeł energii czy prowadzonej polityki energetycznej. Istotny jest również element demografii i preferencji transportowych. Niemniej jednak można zauważyć również, że ocena wpływu na środowisko naturalne w kontekście emisji jest również uwarunkowana kryterium, które jest do takiej oceny wykorzystywane. Stąd tak znaczne czasem różnice pomiędzy danymi krajami i zajmowanie różnych miejsc w rankingach.

Sektor transportu nadal jest uzależniony w dużym stopniu od produktów pochodzących z paliw kopalnianych, takich jak benzyna czy olej napędowy. Rozwiązanie problemu emisji z tych różnych kategorii pojazdów będzie miało kluczowe znaczenie dla osiągnięcia przez UE celów klimatycznych dla sektora transportu. Aby państwa spełniły swoje zobowiązania klimatyczne na rok 2030 i 2040 oraz osiągnęły neutralność klimatyczną, niezbędne jest kontrolowanie poziomów emisyjności gazów cieplarnianych oraz poszukiwanie rozwiązań, które pozwolą poprawić sytuację globalnie, jak i jednostkowo. Aby ograniczyć emisje pojazdów o 94% w stosunku do roku 2005 potrzebne jest znalezienie radykalnych rozwiązań mających na celu ograniczenie emitowanego CO₂ do atmosfery.

Rosnąca liczba pojazdów samochodowych wprowadza konieczność ciągłego podnoszenia standardów w dziedzinie motoryzacji. Wpływ ma na to również ciągły postęp obserwowany w technologiach bezprzewodowych, komunikacyjnych oraz technikach obliczeniowych stosowanych w pojazdach. Równolegle wzrost liczby pojazdów związany jest z powstaniem nowych zagrożeń dla ludzi i środowiska. W związku z tym pożądaną staje się poszukiwanie skutecznych praktyk i rozwiązań w zakresie zwiększania bezpieczeństwa oraz

przeciwdziałania negatywnemu wpływowi na przyrodę. W tym kontekście kluczowa może być analiza dostępnych informacji o stanie pojazdu [41]. Uzyskana w ten sposób wiedza może okazać się przydatna pod kątem badania zachowania kierowcy i jego wpływu na środowisko naturalne i bezpieczeństwo w ruchu drogowym. Potrzeba poszukiwania rozwiązań, które mogłyby przyczynić się do zmniejszenia emisyjności dwutlenku węgla, a tym samym zmniejszenia jego szkodliwego oddziaływania na środowisko jest niepodważalna.

Działania ograniczające emisyjność gazów GHG są głównie skupione na zmniejszeniu emisji CO₂ do środowiska. Przykładem jest zawarty w 1997 roku w Kioto w Japonii międzynarodowy traktat, tzw. protokół Kioto, którego głównym celem było zobowiązanie krajów rozwiniętych do redukcji emisji wybranych gazów cieplarnianych, tj. dwutlenku węgla, metanu i podtlenku azotu. Ustalenia zawarte w tym protokole spotkały się jednak z powszechną krytyką ze względu na brak zobowiązań względem krajów rozwijających się. Kolejnym krokiem w kierunku zwalczania zmian klimatu było przyjęte w 2015 roku w Paryżu, tzw. porozumienie paryskie, które obejmowało zobowiązania dotyczące emisji gazów zarówno ze strony krajów rozwiniętych, jak i rozwijających się. Kluczowym celem tego porozumienia było ustalenie dotyczące utrzymania wzrostu średniej temperatury na świecie poniżej 2°C.

Potrzeba skuteczniejszego włączania się do globalnych wysiłków mających na celu redukcję emisji gazów cieplarnianych wynika przede wszystkim z faktu, iż wiele krajów i regionów w UE wprowadza coraz bardziej rygorystyczne normy emisji spalin. Strategia Unii Europejskiej dotycząca zmniejszenia emisji CO₂ z pojazdów osobowych opiera się na trzech głównych filarach, które stanowi:

- dobrowolne porozumienie z producentami pojazdów w sprawie zmniejszenia emisji CO₂ w nowo produkowanych pojazdach,
- dyrektywa dotycząca oznakowania pojazdów pod kątem emisji CO₂,
- propozycja harmonizacji instrumentów fiskalnych dotyczących samochodów osobowych we wszystkich krajach członkowskich.

Pierwsze dwa filary są obecnie wdrożone i egzekwowane od 1999 r., natomiast najtrudniejszy do osiągnięcia cel stanowi filar dotyczący polityki podatkowej, co wynika z zagwarantowanej suwerenności poszczególnych państw członkowskich. Obecnie coraz częściej kraje UE wprowadzają podatek od emisji CO₂, pobierany w momencie rejestracji nabytego pojazdu i/lub w formie płaconego co roku podatku drogowego. Innym sposobem, praktykowanym przez kraje UE jest wprowadzenie podatków drogowych, nakładanych na branżę transportową w postaci tzw. myta. Wprowadzają one nowy składnik opłaty drogowej, który będzie pobierany w celu odzyskania kosztów spowodowanych emisją CO₂ w trakcie

ruchu pojazdów na drogach płatnych. Dla przykładu takie rozwiązanie funkcjonuje od 1 marca 2024 w Czechach [115]. Kraje Unii Europejskiej powinny współpracować ze sobą w celu wypracowania wspólnej strategii promującej zmniejszenie emisji CO₂. Wśród takich działań nie powinno zabraknąć:

- inwestowania w odnawialne źródła energii, w celu pozyskiwania energii,
- modernizowania transportu, przez rozbudowę sieci transportu publicznego,
- promowania „zielonej logistyki”, zachęcając do korzystania z mniej emisyjnych form transportu,
- finansowania badań nad nowymi technologiami związanymi z możliwościami pozyskiwania i magazynowania CO₂ oraz dalszymi możliwościami jego wykorzystania,
- edukowania i zwiększania świadomości ekologicznej społeczeństwa na temat zmian klimatycznych i ich konsekwencji,
- zazielenienia gospodarki.

Wśród działań związanych z sektorem transportu ważne jest niewątpliwie również zachęcanie do ekonomicznej i ekologicznej jazdy kierowców – tylko jednostkowe podejście może wpłynąć skutecznie na efekt globalny.

2.2. Przegląd metod modelowania emisyjności pojazdów samochodowych

Od momentu pojawienia się na rynku pierwszych samochodów na początku XX wieku do chwili obecnej, pojazdy przeszły ogromną metamorfozę pod kątem właściwie każdego aspektu począwszy od wyglądu, rozmiaru, ciężaru, aż po standardy dotyczące bezpieczeństwa, jakości, ochrony środowiska czy posiadanych funkcji. Niewątpliwy rozwój technologii, z jednej strony ogromnie rozwinął branżę motoryzacyjną, z drugiej jednak doprowadził do omawianych w pracy powszechnie występujących problemów z emisyjnością substancji toksycznych. Należy podkreślić również, że rozwój technologii skutkował wprowadzeniem do pojazdów różnych układów oraz protokołów komunikacyjnych, które zapewniają dostęp do danych generowanych z dostępnych w pojeździe czujników. Dane z układów w przypadku wystąpienia nieprawidłowości mogą być analizowane przez wykwalifikowany personel techniczny jak również i samych kierowców. Ogromna ich ilość oraz dostępność umożliwiają dokładne monitorowanie stanu pojazdu, łącznie z emisją szkodliwych gazów oraz zużycia paliwa.

Kluczowa w kontekście emisyjności pojazdów samochodowych jest również rola kierowcy [16]. Dlatego możliwość diagnostyki stylu jazdy jest ważna szczególnie w celu

wyodrębnienia i docelowo wyeliminowania „negatywnych” oraz „pozytywnych” zachowań kierowcy. Gromadzenie, analiza i ocena takich informacji może pozwolić na odnalezienie wzorców i określenie wpływu wybranych czynników na rozpatrywaną emisyjność pojazdów mechanicznych.

Tworzone są w tym celu modele, które stanowią uproszczony obraz rzeczywistości, służąc do badania i poznawania złożonych układów, zjawisk lub procesów. W celu budowy modeli należy dokonać dokładnej analizy wybranego zjawiska oraz określić czynniki które mogą mieć na nie wpływ. Konstruowanie modeli pozwalających na szacowanie emisji CO₂ z pojazdów samochodowych stało się obszarem o ogromnym znaczeniu, szczególnie na polu naukowym [12]. Rozwijane są w tym kontekście różne metody badawcze, odmienne pod względem złożoności, dokładności, zakresu oraz możliwości zastosowania. Tradycyjne podejścia, takie jak modele analityczne są aktualnie stosowane w połączeniu na przykład z technikami wykorzystującymi sztuczną inteligencję, która oferuje możliwość modelowania emisyjności w oparciu o złożone algorytmy uczenia maszynowego. Emisja CO₂ w pojazdach samochodowych jest złożonym problemem, zależnym od wielu czynników, które na nią wpływają. Ich poprawne zdefiniowanie jest kluczowe w kontekście opracowywania i wdrażania strategii redukcji emisji CO₂, a szczególnie w celu zaspokajania potrzeby głębszego zrozumienia, analizy i oceny wpływu poszczególnych zmiennych na emisję CO₂. Dlatego - na podstawie analizy literatury - wyszczególniono najważniejsze z nich, do których zaliczono:

- rodzaj paliwa, którym zasilany jest pojazd [95],
- metody pomiarów poziomu emisji CO₂ [37],
- styl jazdy kierowcy [6], [28], [34], [121],
- ilość zużytego paliwa [34], [37], [77], [95],
- udział odbiorników energii elektrycznej w pojazdach wpływających na komfort podróżowania i właściwości jezdne pojazdu, tzw. systemów funkcjonalnych [77],
- wartości kinematyczne charakteryzujące ruch pojazdu, tj. prędkość, przyspieszenie pojazdu [34], [121], [95],
- zastosowane systemy diagnostyczne [42], [75].

Silniki spalinowe współczesnych samochodów zasilane są głównie benzyną lub olejem napędowym. Analizę danych dotyczących emisji CO₂ i zużycia paliwa w wybranych spalinowych pojazdach samochodowych można odnaleźć między innymi w artykule [88]. Opisano w nim również różnice w regulacjach dotyczących emisji CO₂ w różnych krajach, a następnie, na podstawie zebranych danych dotyczących emisji CO₂ i zużycia paliwa, dokonano ich porównania z teoretycznymi obliczeniami stechiometrycznymi dla emisji CO₂

w reakcji zużycia paliwa. Wykorzystano w tym celu model regresji liniowej, a otrzymane wyniki pozwoliły ustalić, iż silniki o zapłonie samoczynnym, pomimo niższego zużycia paliwa, emitują więcej CO₂ na jednostkę przebytego dystansu w stosunku do silników benzynowych. Ponadto wykazano wysoki współczynnik korelacji wynoszący ok. 0,999 dla zmiennych emisja CO₂ oraz zużycie paliwa dla obydwu wymienionych rodzajów silników oraz istnienie zależności liniowej pomiędzy emisją CO₂, a zużyciem paliwa.

Problematyka dostosowywania się producentów pojazdów do obowiązujących regulacji dotyczących norm emisji CO₂ jest uzależniona od możliwości dokładnego pomiaru, m.in. poziomu zużycia paliwa oraz emisji CO₂ w warunkach przeciętnej eksploatacji pojazdu. Z uwagi na konieczność zapewnienia powtarzalności wyników, wspomnianych pomiarów dokonuje się w warunkach laboratoryjnych, co może prowadzić do występowania wyraźnych odstępstw pomiędzy pomiarami laboratoryjnymi i tymi podczas rzeczywistego użytkowania samochodu. W artykule [37] dokonano przeglądu przyczyn różnic pomiędzy zmierzonym zużyciem paliwa w warunkach laboratoryjnych, a rzeczywistym oraz emisją CO₂ z samochodów osobowych wykorzystywanych w Europie. Pomimo odnotowanego, znaczącego spadku sumarycznej emisji całkowitej wykazano, że rzeczywiste wartości emisji są często wyższe niż te uzyskane w warunkach laboratoryjnych, co zmusza producentów pojazdów do weryfikacji obowiązkowych celów w zakresie redukcji CO₂. W artykule tym ponadto pogrupowano czynniki wpływające na zużycie paliwa i emisję CO₂ na trzy główne obszary: pojazd, warunki środowiskowe oraz kierowca.

Kierowca jest kolejnym czynnikiem, który należy wziąć pod uwagę podczas analizy emisji CO₂. W artykule [6] badano wpływ stylu jazdy kierowcy na emisję CO₂ i zużycie paliwa w transporcie drogowym wykazując, że ma on istotne znaczenie. Stwierdzone różnice, określone jako „spokojna” i „agresywna” jazda, mogą prowadzić nawet do 40% różnic w zużyciu paliwa i w poziomie emisji CO₂. Wykorzystanie podczas jazdy techniki Eco-drivingu, polegającej na przewidywaniu sytuacji na drodze, unikaniu zbędnych przyspieszeń i hamowania oraz efektywnym wykorzystaniu silnika może przynieść wymierne rezultaty. Wykazano to za pomocą opracowanego algorytmu Ecoindex [6], służącego do określenia wskaźnika zużycia paliwa oraz emisji CO₂, który wynika z zachowania kierowcy w zależności od innych czynników, tj. trasy pojazdu czy warunków na drodze. Algorytm dokonuje kalkulacji zużycia paliwa i emisji CO₂ w zmodyfikowanym cyklu jazdy zakładając, iż kierowca przestrzegał zasad ekologicznej jazdy. Na podstawie wyników badań stwierdzono, iż przy niższych prędkościach ruchu pojazdu, w warunkach miejskich, zachowanie kierowcy ma większy wpływ na zużycie paliwa i emisję CO₂ w porównaniu z jazdą po autostradzie, gdzie

wpływ ten jest znacznie mniejszy. Natomiast autorzy artykułu [34] zwrócili uwagę, że na wzrost zużycia paliwa oraz emisji CO₂ istotny wpływ ma również „agresywny” styl jazdy, charakteryzujący się dużymi prędkościami pojazdu, zwłaszcza tymi przekraczającymi 100 km/h. Z kolei w pracy [79] wykazano, iż na zmniejszenie emisyjności wpływa także minimalizacja czasu jazdy na niskich biegach i jak najszybsze przechodzenie na wyższe, przy jednoczesnym ograniczeniu ich częstej zmiany (w przypadku manualnej skrzyni biegów).

Kierowca, wraz ze wszystkimi podejmowanymi przez siebie w trakcie jazdy decyzjami, ma zatem kluczowy wpływ na emisję CO₂, jednak z uwagi na występujący ruch innych kierowców na drodze i czynniki zewnętrzne nie zawsze może poruszać się w sposób zrównoważony i ze stałą prędkością. W tym kontekście, w artykule [28], zbadano zmiany w emisji CO₂ w przypadku pojazdów poruszających się ze zmienną prędkością. Jest to niezwykle istotne z uwagi na wpływ wahań prędkości w fazie przyspieszania na wydatki energetyczne i emisję gazów cieplarnianych. Przeprowadzone badania miały na celu umożliwienie wyprowadzenia teoretycznego modelu emisji CO₂ w oparciu o podstawowe prawa fizyki. Wyniki wykazały, że maksymalne błędy względne prognoz emisji CO₂ dla samochodów osobowych i ciężarowych wyniosły odpowiednio 9,75% i 10,10%.

Po uwzględnieniu wpływu kierowcy należy również wziąć pod uwagę oddziaływanie systemów funkcjonalnych, których głównym zadaniem jest zapewnienie lepszych właściwości jezdnych pojazdu i zwiększenie komfortu podróżowania pasażerów. W artykule [77] zbadano, w jaki sposób wpływają one na zużycie paliwa i emisję CO₂. Praca systemów funkcjonalnych wiąże się z dodatkowym poborem energii elektrycznej i stanowi źródło obciążenia silnika. Przykładami tego typu systemów są m. in. układ klimatyzacji pojazdu, wentylowane/podgrzewane fotele, oświetlenie wnętrza samochodu, system audiowizualny, tryb jazdy pojazdu itd. Wspomniane systemy zostały poddane ocenie wpływu na poziom emisji gazów cieplarnianych w rzeczywistych warunkach ruchu drogowego przy użyciu mobilnego systemu pomiaru emisji (PEMS). Z przeprowadzonych w artykule badań wynika, że stosowanie takich udogodnień ma znaczny wpływ zarówno na emisję jak i zużycie paliwa, powodując ich wzrost, który w niektórych przypadkach przekracza 50%. Podkreśla to potrzebę uwzględnienia ekologicznych i ekonomicznych konsekwencji wdrażania tych systemów zarówno w projektowaniu jak i eksploatacji pojazdów. Wskazuje również potrzebę dalszych badań w tym obszarze, mając szczególnie na uwadze aktualne trendy w wyposażeniu pojazdów.

Kolejnym czynnikiem, mającym wpływ na emisję CO₂, jest prędkość pojazdu oraz jego przyspieszenie. Wspomnianą tematyką zajęli się między innymi autorzy artykułu [81], w którym skupiono się na opracowaniu modeli przewidywania emisji dwutlenku węgla dla

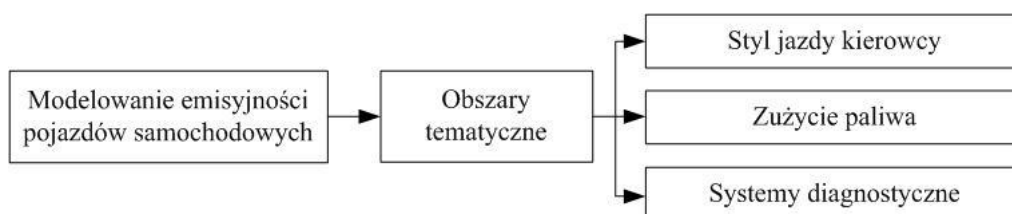
wartości chwilowej prędkości i przyspieszenia pojazdów, stosując w tym celu metodę regresji matematycznej. Stwierdzono istotną, liniową zależność pomiędzy emisją CO₂, a prędkością oraz zmianą przyspieszania pojazdu. Wyniki podkreślają konieczność uwzględniania zmian zarówno prędkości jak i przyspieszenia w modelach emisji CO₂, w celu uzyskania dokładnych prognoz.

Kierowca w trakcie jazdy może mieć dostęp do informacji dotyczących chwilowego zużycia paliwa oraz poziomu emisji CO₂, pochodzących z pokładowego systemu diagnostyki samochodowej. Wskazania te mogą być jednak obarczone pewnymi poziomami błędów dotyczących dokładności pomiaru zużycia paliwa i emisji [8]. Badania emisji NO₂ i CO₂ w odniesieniu do pojazdów ciężarowych z silnikami o zapłonie samoczynnym za pomocą systemów diagnostyki pokładowej badano w artykule [42]. W wymienionej pracy wykorzystano dane pochodzące właśnie z systemu diagnostyki samochodowej (OBD-II), które następnie zestawiono z eksperymentalnymi danymi, uzyskanymi przy pomocy zewnętrznej aparatury pomiarowej, uzyskanymi na hamowni w cyklu jazdy CHTC-HT (z ang. China Automotive Test Cycle of Heavy-Duty Trucks). Autorzy udowodnili, iż wyniki obliczone na podstawie danych uzyskanych z systemu diagnostyki są zgodne z wynikami mierzonymi przez analizator emisji, ponieważ emisje CO₂ są ściśle związane ze zużyciem paliwa przez silnik pojazdu. Emisje CO₂ mierzone przez analizator emisji miały natomiast mniejsze chwilowe wahania niż wskazania na podstawie danych OBD-II, co można tłumaczyć m. in. ruchem spalin w rurze wylotowej.

Ciekawych wyników dostarcza połączenie funkcjonalne, które występuje w tzw. hybrydach, pojazdach złożonych z dwóch typów silników napędów, np. spalinowego i elektrycznego. Koegzystencja dwóch różnych silników w ramach jednego pojazdu stanowi kolejny czynnik wpływający na emisję CO₂. Stanowiło to przedmiot badań w pracy [75], w której wykorzystano techniki uczenia maszynowego opierając się na danych z mobilnego systemu pomiaru emisji (PEMS) oraz z systemu OBD-II. Uzyskano różnice w wynikach emisji z systemu PEMS z badań drogowych i obliczeń modelowych, przy czym największe rozbieżności dotyczyły odcinków dróg wiejskich. Opracowany model wymagał podania danych dotyczących prędkości, przyspieszenia i nachylenia drogi, które uzyskano z testów drogowych. Ponadto z wyników badań wynika, że emisja CO₂ w przypadku pojazdów hybrydowych nie jest ściśle powiązana z prędkością i przyspieszeniem pojazdu ze względu na okresy, w których silnik spalinowy nie pracuje.

Podsumowując, problem emisyjności pojazdów samochodowych jest problemem złożonym. Na podstawie literatury przedmiotu można stwierdzić, że czynnikami wpływającymi

na emisyjność pojazdów samochodach definiowaną na poziomie pojedynczego użytkownika są: styl jazdy kierowców, zużycie paliwa oraz systemy diagnostyczne. Pozostałe charakteryzowane czynniki są natomiast z nimi powiązane. Styl jazdy kierowcy determinuje kinematyczne parametry ruchu pojazdu, normy i metody pomiarów poziomów emisji CO₂ są ściśle związane z dostępnymi, pokładowymi systemami diagnostycznymi, a systemy funkcjonalne z kolei dotyczą procesu zużycia paliwa itd. Wskazano tym samym trzy kluczowe obszary tematyczne dotyczące problematyki modelowania emisyjności pojazdów samochodowych, co zostało przedstawione schematycznie na rysunku 20.



Rys. 20. Kluczowe obszary badań w zakresie emisyjności pojazdów

Wymienione obszary zostaną szczegółowo omówione w dalszych częściach pracy.

2.3. Styl jazdy kierowcy w ujęciu literaturowym

Na ruch pojazdu kluczowy wpływ ma kierowca i jego zamiar (plan jazdy), a samochód jest narzędziem, które pozwala go zrealizować. Każdy kierowca działa w indywidualny sposób, jednak analiza dużej grupy wykazała występowanie pewnych powtarzalnych schematów w ich zachowaniu, które klasyfikuje się jako tzw. styl jazdy, stanowiący pierwszy z wymienionych obszarów tematycznych (rys. 20) wykazanych w poprzednim podrozdziale [94]. W literaturze wyróżnia się kilka odrębnych stylów jazdy, wśród których każdy z nich ma indywidualne cechy i określony wpływ na emisję CO₂. Występuje wiele różnych podziałów, np. autorzy artykułu [104] wprowadzają podział na style: agresywny, normalny, niebezpieczny. Najczęściej jednak wyróżnia się trzy style jazdy: agresywny, zrównoważony oraz spokojny [9]. Style: spokojny i agresywny stanowią dwie antagonistycznie filozofie prowadzenia pojazdu, skupiające się na przeciwstawnych dążeniach w kontekście, m.in. osiąganych prędkości i przyspieszeń (opóźnień) pojazdu, liczby wykonywanych manewrów itd. Spokojny styl jazdy zakłada minimalizację wartości parametrów kinematycznych pojazdu, tj. prędkości, przyspieszenia oraz minimalizację liczby wykonywanych manewrów podczas poruszania się. Kierowcy tego typu mają tendencję do poruszania się z mniejszą niż maksymalna dopuszczalna prędkość w danym obszarze, powolnego przyspieszania oraz niewykazywania skłonności do gwałtownych zachowań na drodze. Z kolei agresywny styl jazdy kierowców zakłada przeciwstawne tendencje, tzn. obejmuje jazdę z wysoką prędkością oraz gwałtowne

zachowania, tj. częste nagłe przyspieszanie, hamowanie, wyprzedzanie itd. Oba wymienione style mogą prowadzić do zwiększonego zużycia paliwa, które przekłada się na również zwiększoną emisję CO₂. W przypadku spokojnego stylu jazdy wynika to z poruszania się z często zbyt niskimi wartościami przełożeń silnika, natomiast w przypadku agresywnego stylu jazdy ze zbyt dużych przyspieszeń i prędkości pojazdu. Zrównoważony styl jazdy stanowi natomiast sposób prowadzenia pojazdu umożliwiający zmniejszenie emisji CO₂ generowanej przez samochód i wynika ze świadomych decyzji podejmowanych przez kierowcę w celu zwiększenia wydajności pracy silnika. Stanowi on kompromis pomiędzy spokojnym i agresywnym stylem jazdy, charakteryzując się płynnym przyspieszaniem i hamowaniem, utrzymywaniem stałej prędkości pojazdu przy jednoczesnej minimalizacji liczby wykonywanych niezbędnych manewrów. Wszystkie opisane decyzje stanowią wielowymiarowy problem i oprócz samego stylu jazdy obejmują również dobór pojazdu i proces planowania trasy. Modelowanie stylu jazdy kierowców, w celu docelowego określenia w jaki sposób maksymalizować prędkość pojazdu zgodnie z obowiązującymi limitami prędkości i zachowaniem bezpieczeństwa w ruchu drogowym, przy jednoczesnej minimalnej emisji CO₂, może mieć realny wpływ na środowisko naturalne. Dotyczy to nie tylko indywidualnych kierowców, ale również całych firm transportowych, zarządzających flotą samochodów. W związku z tym analiza zachowania kierowców na drodze jest niezwykle ważnym zagadnieniem, gdyż umożliwia wskazanie stylu jazdy charakteryzującego się minimalną emisją CO₂.

Przeprowadzona kwerenda literatury w tym obszarze tematycznym wykazała istnienie dwóch grup czynników: odzwierciedlających oraz wpływających na zachowanie kierowców, które zostały wymienione w tabeli 5.

Tabela 5. Grupy czynników wpływających na zachowanie kierowców na drodze [69]

Grupa	Czynniki	Źródła
Czynniki odzwierciedlające zachowanie kierowców	Prędkość pojazdu	[16], [18], [56], [65], [86], [104], [120]
	Przyspieszenie pojazdu	[16], [56], [65], [86], [104], [120]
	Opóźnienie pojazdu	[56], [86], [120]
	Pedały	[56], [65], [86], [120]
	Kierownica	[56], [86], [120]
	Bieg	[18], [56], [86], [104], [120]
	Silnik	[18], [56], [65], [86], [104], [120]
Czynniki wpływające na zachowanie kierowców	Sytuacja drogowa	[10], [11]
	Warunki pogodowe	[17], [36]

Grupa	Czynniki	Źródła
	Sygnalizacja świetlna	[10], [11]
	Warunki drogowe	[36], [113]
	Cechy psychologiczne	[7],[50], [73]

Do czynników odzwierciedlających zachowanie kierowców można zaliczyć m.in. parametry kinematyczne pojazdu, np. prędkość, przyspieszenie (opóźnienie) pojazdu, położenie pedałów, np. przyspieszenia, hamulca, sprzęgła (w przypadku manualnej skrzyni biegów), kierownicy, aktualne przełożenie skrzyni biegów (wartość zależy od wyboru biegu dokonanego przez kierowcę), parametry silnika itd. Z kolei do czynników wpływających na zachowanie kierowców można zaliczyć panującą na drodze sytuację, warunki pogodowe, położenie i stan sygnalizacji świetlnej, stan infrastruktury drogowej, cechy psychologiczne kierowcy itd. Wymienione powyżej czynniki zostały zdefiniowane szczegółowo w tabeli 6.

Tabela 6. Definicja czynników wpływających na zachowanie kierowców na drodze [69]

Grupa	Czynniki	Definicja
Czynniki odzwierciedlające zachowanie kierowców	Prędkość pojazdu	Wartość chwilowej prędkości liniowej pojazdu
	Przyspieszenie pojazdu	Wartość chwilowego przyspieszenia liniowego pojazdu
	Opóźnienie	Wartość chwilowego opóźnienia liniowego pojazdu
	Pedały	Wartość chwilowego położenia kąтового pedału przyspieszenia, sprzęgła oraz hamulca
	Kierownica	Wartość chwilowego kąta obrotu kierownicy
	Bieg	Wartość aktualnego przełożenia skrzyni biegów (manualnej)
	Silnik	Wartość chwilowej prędkości kątowej silnika, obciążenia silnika, mocy silnika
Czynniki wpływające na zachowanie kierowców	Sytuacja drogowa	Odległości pomiędzy samochodami, odległość pomiędzy samochodem, a sygnalizacją świetlną itd.
	Warunki pogodowe	Temperatura otoczenia, widoczność, opady deszczu, wilgotność, prędkość wiatru
	Sygnalizacja świetlna	Sygnały generowane przez sygnalizację świetlną
	Infrastruktura drogowa	Geometria drogi, nachylenie terenu, krzywizna drogi
	Czynniki psychologiczne	Stres, Doświadczenie w prowadzeniu pojazdów, cechy personalne kierowcy

Grupa pierwszych wymienionych w tabeli 6 czynników stanowi zestaw parametrów mierzalnych, które mogą służyć do analizy zachowania kierowcy na drodze. Z kolei grupa czynników wpływających na zachowanie kierowców zawiera parametry opisujące warunki

pogodowe, sytuację drogową oraz parametry trudno mierzalne i dość skomplikowane w opisie ze względu na swoją nieprzewidywalną naturę, np. stres czy cechy personalne kierowcy.

W przypadku wszystkich omawianych czynników opisujących zachowanie kierowców należy wyróżnić w tym kontekście dwa podstawowe problemy: klasyfikacji i identyfikacji. Klasyfikacja jest zadaniem polegającym na ustaleniu pewnych kategorii lub grup, do których można przypisać i tym samym uporządkować dane badane zjawisko (np. styl jazdy kierowcy). Z kolei zadanie identyfikacji dotyczy nadania konkretnemu zjawisku kategorii klasyfikacyjnej na podstawie rozpatrywanych cech. Klasyfikacja jest zatem bardziej rozległym problemem, dążącym do zdefiniowania kategorii, natomiast identyfikacja polega na określaniu, do której kategorii lub grupy pasuje dane zjawisko.

Klasyfikacja kierowców na różne grupy, np. kierowców „zrównoważonych” i „agresywnych” umożliwia kategoryzowanie ich zachowań na podstawie czynników wymienionych w tabeli 6, zarówno wpływających, jak również odzwierciedlających zachowanie kierowców. Klasyfikacja umożliwia także lepsze zrozumienie wpływu czynników kształtujących zachowanie kierowców. Z kolei identyfikacja poszczególnych grup kierowców stanowi kluczowy aspekt umożliwiający zrozumienie ich zachowań. Skuteczna klasyfikacja i identyfikacja kierowców na podstawie opisywanych czynników ma kluczowe znaczenie dla zrozumienia wpływu stylu jazdy kierowców na emisyjność pojazdów samochodowych. W tabeli 7 przedstawiono przykłady badań naukowych opisujących tematykę związaną z dwoma wymienionymi problemami.

Tabela 7. Grupy podstawowych problemów opisujących zachowanie kierowców

Grupa	Źródła
Problemy klasyfikacyjne	[16], [18], [65], [104]
Problemy dotyczące identyfikacji	[56], [83], [86], [120]

Zachowanie kierowcy może być również identyfikowane za pomocą współczesnych technologii, które można wykorzystać do śledzenia i analizowania czynności wykonywanych za kierownicą. Dynamicznie zmieniające się parametry samochodu, np. prędkość liniowa pojazdu, przyspieszenie liniowe pojazdu, prędkość kątowna wału korbowego silnika mogą być zapisywane za pomocą m.in. współczesnych systemów diagnostyki, opartych o czujniki występujące na pokładzie pojazdów samochodowych (np. monitorujących parametry pracy jednostki napędowej samochodu), systemy GNSS, INS itd. Większość obecnie produkowanych pojazdów samochodowych jest wyposażona w złącza (porty) diagnostyczne pozwalające kierowcy lub technikowi diagnostyce na ocenę stanu technicznego danego układu pojazdu [26], [64], [121]. Stosowane w tym celu skanery OBD są niewielkimi urządzeniami elektronicznymi

służącymi do gromadzenia informacji [18], [43], [72], [117] (tematyka ta zostanie omówiona szczegółowo w podrozdziale 2.4). Ich coraz powszechniejsze stosowanie i rosnąca dostępność gromadzonych danych, mogą stanowić podstawę do tworzenia różnych modeli matematycznych.

Sytuacja drogowa jest jednym z czynników wpływających na zachowanie kierowców. Parametrem opisującym wymieniony czynnik jest odległość pomiędzy samochodem a innymi poruszającymi się pojazdami w ruchu drogowym lub sygnalizacją świetlną, znakami drogowymi itd. [10], [11]. Występujące odległości wymuszają na kierowcy konieczność dostosowania prędkości pojazdu do aktualnie panujących warunków drogowych. Kierowcy muszą stale monitorować wspomniane odległości, aby zachować bezpieczny dystans, który zapewni im wystarczający czas na reakcję w przypadku konieczności nagłego hamowania lub zmiany pasa ruchu.

Obecność sygnalizacji świetlnej stanowi również czynnik mający wpływ na zachowanie kierowców, gdyż nagłe zmiany kolorów wymuszają podjęcie określonych działań jak np. użycie pedałów hamulca oraz sprzęgła i zmiany przełożenia skrzyni biegów (w przypadku manualnej skrzyni biegów) celem ograniczenia prędkości liniowej pojazdu [10], [11]. Z kolei niezastosowanie się do sygnałów świetlnych może skutkować wykroczeniami drogowymi, a nawet wypadkami i powodować konsekwencje prawne.

Kolejnym, niezwykle istotnym czynnikiem są warunki pogodowe, a szczególnie temperatura otoczenia, wilgotność, opady deszczu, prędkość wiatru, które znacząco wpływają na styl jazdy kierowców. Nagłe zmiany temperatury i wilgotności wpływają na współczynnik oporu ruchu i w rezultacie zależność pomiędzy wartością przyspieszenia liniowego pojazdu a chwilowym położeniem kątowym pedału przyspieszenia ulega zmianie, zmuszając tym samym kierowcę do ograniczania prędkości liniowej pojazdu [17], [36]. Parametry opisujące warunki pogodowe, tj. temperatura, wilgotność, prędkość wiatru są możliwe do pomiaru wyłącznie w oparciu o czujniki zewnętrzne.

Na zachowanie kierowców ma wpływ również stan infrastruktury drogowej, do której zalicza się m.in. geometrię drogi, nachylenie terenu czy krzywiznę drogi [36], [113]. Istotna jest m.in. liczba pasów ruchu czy ich szerokość, gdyż może mieć wpływ na sposób poruszania się i manewrowania pojazdami przez kierowców. Z kolei nachylenie drogi ma znaczący wpływ na zachowanie kierowcy, szczególnie podczas podjazdu lub zjazdu ze stromych wzniesień. Kierowca w takim wypadku musi zredukować aktualne przełożenie skrzyni biegów na niższe (w przypadku manualnej skrzyni biegów) oraz dostosować prędkość liniową pojazdu do występujących warunków. Krzywizna drogi również wpływa na zachowanie kierowcy,

ponieważ musi on dostosować prędkość i położenie pojazdu do aktualnie zajmowanego pasa ruchu, aby bezpiecznie pokonywać zakręty. Bardziej złożone zakręty wymagają od kierowców użycia pedału hamulca i zwolnienia, natomiast łagodniejsze mogą pozwolić na osiągnięcie wyższych prędkości.

Ostatnim czynnikiem wymienionym w tabeli 6 są cechy psychologiczne kierowców, analizowane często również w kontekście ekologii. Za przykład można podać artykuł [50], w którym zbadano wpływ wybranych czynników ludzkich, tj. doświadczenia w prowadzeniu samochodów czy poziomu stresu, które znacząco wpływają na zachowanie kierowcy, jednak są trudno mierzalne. W artykule zastosowano techniki badań ankietowych na grupie 59 kierowców. Wykorzystanie wyników badań podczas kształcenia przyszłych kierowców może wpłynąć także na zmniejszenie emisji CO₂ oraz zużycia paliwa przez promowanie „zrównoważonego” stylu jazdy. Kolejnymi ważnymi czynnikami psychologicznymi wpływającymi na zachowanie kierowców są cechy personalne/osobowościowe kierowcy. W artykule [73] zwrócono uwagę na wpływ „agresywnych” zachowań kierowcy na zużycie paliwa (również na emisję CO₂), tj. gwałtowne reakcje na zdarzenia drogowe, nagłe przyspieszanie oraz hamowanie, natychmiastowe ruszanie z miejsca itd. Autorzy zaproponowali metodę profilowania w celu zachęcenia do bardziej oszczędnej jazdy. Wspomniana metoda skupia się na wpływaniu na użytkowników i doskonaleniu ich zachowań przez grę. W przypadku wykrycia sytuacji niekorzystnej, scenariusz gry proponuje kierowcy kilka opcji do wyboru, natomiast kierowca może następnie zmodyfikować sposób prowadzenia pojazdu, aby zmniejszyć zużycie paliwa i emisję CO₂. Wyniki bazują na analizie następujących parametrów, tj. prędkość i przyspieszenie pojazdu, prędkość obrotowa silnika, czy położenie przepustnicy (wyłącznie dla silników ZI). Z kolei samym celem gry jest określenie poziomu „agresji” kierowcy, która bezpośrednio przekłada się na zwiększone zużycie paliwa. Gra przyznaje kierowcy punktację po przejeździe w zakresie od 0 (najgorszy wynik) do 100 (najlepszy wynik) i na tej podstawie dokonuje jego klasyfikacji. Z kolei autorzy artykułu [7] skupili się na demograficznych różnicach pomiędzy różnymi grupami kierowców, tj. kierowcami płci męskiej i żeńskiej, młodszymi i starszymi itd. Dokonany przegląd badań zachowań kierowców opartych na przetwarzaniu danych z czujników wykazał, iż w ok. 80% badań bierze udział większa liczba mężczyzn niż kobiet. Autorzy wskazują na problem z brakiem uwzględnienia zróżnicowania płci w testach. Autorzy zaprezentowali również wytyczne, oparte na przeglądzie analizy stanu zagadnienia, dotyczące najskuteczniejszego sposobu profilowania kierowców na podstawie ich zachowania.

Kolejnym zagadnieniem, związanym z wpływem kierowców na emisję CO₂, jest problematyka identyfikacji kierowców [100]. W tym celu wykorzystuje się różne metody, na przykład regresję logistyczną. Autorzy na podstawie danych z systemu OBD-II, skupili się na identyfikacji niebezpiecznych zachowań kierowców przez analizę wybranych czynników, tj. liczby gwałtownych hamowań, wartości maksymalnej prędkości pojazdu oraz całkowitego czasu jazdy. Opracowany przez autorów model służy, m. in. do oceny ryzyka niebezpiecznej jazdy. Wyniki badań dostarczyły ważnych informacji o wpływie różnych czynników na bezpieczne zachowania kierowców.

W celu identyfikacji kierowców można również zastosować bardziej złożone algorytmy uczenia maszynowego, tj. SVM, czy drzewa decyzyjne. Przykładem takich prac może być artykuł [56], w którym autorzy dokonali identyfikacji kierowców na podstawie danych z systemu OBD-II. Do analizy danych i identyfikacji kierowców wykorzystali takie metody jak: k-NN, SVM, regresja logistyczna oraz drzewa decyzyjne. Wyniki badań pokazują, że identyfikacja kierowców jest możliwa z zadowalającą dokładnością, choć dokładność ta zależy od liczby kierowców w zestawie danych. W badaniu autorzy przeanalizowali 51 cech, tj.: prędkość pojazdu, przyspieszenie pojazdu, prędkość obrotową silnika, temperaturę powietrza dolotowego, temperaturę chłodziwa, pozycję przepustnicy, masowy przepływ powietrza, napięcie sondy lambda, czas włączenia silnika, ilość paliwa użytego od ostatniego resetu licznika, czas pracy silnika przy wysokich obrotach, czas pracy silnika przy niskich obrotach, średnią prędkość, maksymalną prędkość i całkowitą przejechaną odległość itd.

Kolejnym przykładem badań dotyczących identyfikacji kierowcy jest artykuł [120]. Autorzy na podstawie publicznie dostępnych danych z systemu OBD-II pojazdów opracowali modele identyfikacji kierowców z wykorzystaniem algorytmów uczenia maszynowego, tj.: k-NN, drzew decyzyjnych, lasów losowych, wzmocnień gradientowych i SVM. Do budowy wspomnianych modeli autorzy przeanalizowali dane, tj.: prędkość obrotową silnika, prędkość pojazdu, temperaturę powietrza dolotowego, położenie pedału przyspieszenia, pozycję przepustnicy, moment obrotowy silnika oraz temperaturę cieczy chłodzącej silnika. Celem badania była identyfikacja kierowców na podstawie ich stylu jazdy.

Do identyfikacji zachowań kierowców można również wykorzystać głębokie sieci neuronowe. W artykule [86] dokonano oceny wybranych metod uczenia maszynowego na potrzeby analizy zachowań kierowców w przypadku dysponowania dużą ilością danych z pojazdów obejmujących, m.in. prędkość pojazdu, przyspieszenie pojazdu, zużycie paliwa, kierunek ruchu itd. Wyniki pokazały, iż wspomniane metody mogą być z powodzeniem stosowane, jednak w celu umożliwienia wydajnej pracy obydwu metod należy zastosować

architekturę opartą o chmurę obliczeniową z uwagi na zapotrzebowanie na dużą moc obliczeniową.

Kolejnym problemem związanym z omawianą tematyką jest klasyfikacja kierowców, do której również można – oprócz takich metod uczenia maszynowego jak SVM, lasy losowe – zastosować także metody boostingu, tj. algorytmy AdaBoost. Przykładem pracy wykorzystującej wspomnianą metodę jest artykuł [65], w którym skupiono się na analizie zachowań kierowców z wykorzystaniem danych z OBD-II oraz wybranych metod uczenia maszynowego. Badanie wykorzystuje wymienione algorytmy (SVM, AdaBoost i lasy losowe) do klasyfikacji zachowań kierowców na podstawie ponad 50 zmiennych. Wybrano zużycie paliwa, pozycję pedału przyspieszenia, prędkość obrotową silnika, ciśnienie powietrza w kolektorze dolotowym, położenie pedału przyspieszenia, bezwzględne położenie przepustnicy, moment obrotowy silnika, obliczoną wartość obciążenia, minimalny i maksymalny wskazany moment obrotowy silnika, moment bezwładności koła zamachowego, aktualny bieg, temperatura płynu chłodzącego, kąt skrętu kierownicy, czas, itd. Autorzy utworzyli i zweryfikowali model kategoryzowania zachowań kierowców za pomocą algorytmów SVM, ADABOOST i lasów losowych. Zaproponowano dziesięć klas odpowiednich zachowań podczas jazdy, które w systematyczny i logiczny sposób prezentują na jakie działania kierowcy mają bezpośredni i pośredni wpływ. Model klasyfikacji zachowań stworzony w ramach tego badania pozwala na ocenę, w jaki sposób działania kierowcy mogą mieć bezpośredni lub pośredni wpływ na zużycie paliwa oraz emisję CO₂. Zaproponowane metody osiągnęły dokładność, która wynosiła ok. 99% (SVM i ADABOOST) i 100% (las losowe). Badanie wykazało wysoką skuteczność tych metod w klasyfikacji zachowań kierowców, co może mieć zastosowanie w poprawie bezpieczeństwa drogowego i efektywności jazdy.

Podobną tematyką zajęli się autorzy artykułu [18], w którym, w celu analizy zachowania kierowcy, zastosowali algorytm AdaBoost w oparciu o dane z pojazdu uzyskane z OBD-II, tj. prędkość samochodu, prędkość obrotowa silnika, położenie przepustnicy i obciążenie silnika. Proponowana metoda, wykorzystująca algorytm AdaBoost, pozwala uzyskać 100% dokładności klasyfikacji już w około 15 iteracjach. Autorzy wskazali, iż proponowaną technikę można zastosować w ramach systemu wspomagania kierowcy.

Klasyfikacja stylów jazdy była również tematem artykułu [104], w którym została przedstawiona możliwość wykorzystania danych uzyskanych przy pomocy interfejsu OBD-II do analizy wzorców jazdy kierowców za pomocą algorytmów uczenia maszynowego. Dane z systemu były zbierane i przesyłane do chmury, a następnie analizowane z wykorzystaniem sztucznej inteligencji, w celu identyfikacji różnych stylów jazdy

(agresywny, normalny, niebezpieczny). Celem badania była poprawa bezpieczeństwa drogowego przez identyfikację i korygowanie niebezpiecznych zachowań kierowców. W artykule autorzy wykorzystali analizę skupień i algorytmy uczenia maszynowego do grupowania i identyfikacji różnych stylów jazdy kierowców. Do budowy modelu wykorzystali parametry kinematyczne z systemu OBD-II, takie jak prędkość pojazdu, przyspieszenie pojazdu, hamowanie i styl jazdy.

Niektórzy badacze, zamiast dokonywać klasyfikacji podstawowych stylów jazdy, skupili się wyłącznie na ich ekologicznym aspekcie. W artykule [16] utworzono model oceny zachowań ekologicznych kierowcy na podstawie danych eksploatacyjnych pojazdu i zużycia paliwa. Dokonano badania trzech charakterystycznych parametrów: prędkość pojazdu, tryb jazdy i zużycie paliwa. Model oceny ekologicznej jazdy stworzono przy użyciu analizy głównych składowych (PCA) i wielokrotnej regresji liniowej. Podstawą tego modelu jest ilościowa ocena wpływu zachowania kierowcy na zużycie paliwa oraz kilka praktycznych instrukcji opartych na zdarzeniach związanych z wspomnianą jazdą. Wyniki badań można bezpośrednio zastosować do udoskonalenia systemów wspomagających ekologiczną jazdę.

Kolejnym tematem, który należy wziąć pod uwagę w analizie stylu jazdy kierowcy, są metody pozyskiwania danych do modelowania stylów jazdy kierowców. W artykule [64] dokonano badań z wykorzystaniem skanera OBD-II wraz z modułem Bluetooth oraz oprogramowania kompatybilnego z protokołem OBD-II w celu akwizycji danych dotyczących stanu pojazdu. Wzięto pod uwagę następujące parametry: prędkość pojazdu, prędkość obrotowa silnika, położenie przepustnicy, które następnie wyeksportowano i zbadano podczas eksperymentu. Z uwagi na występujące opóźnienia w komunikacji bezprzewodowej w otrzymywanych danych ze skanera, zdecydowano się minimalizować liczbę parametrów. Ponadto z samego ECU (z ang. Electronic Control Unit - jednostka sterująca silnika) można zapisać około 50 parametrów, które mogą posłużyć m. in. do określenia stylu jazdy kierowcy, diagnostyki pojazdu, planowania przeglądów, napraw itd.

Oprócz skanerów OBD-II można również zastosować powszechnie dostępne telefony komórkowe wraz ze specjalną przystawką oraz inne czujniki, które wraz z dedykowanymi aplikacjami mogą śledzić zachowania kierowców, co było przedmiotem badania w wielu publikacjach naukowych [15], [55], [57], [64], [72], [79]. Z kolei autorzy artykułu [102] zaprezentowali opracowaną aplikację na telefon komórkowy z systemem Android, która – oprócz wspomnianego śledzenia zachowania kierowców - ostrzega również o wszelkich nieprawidłowościach w ich stylu jazdy. Aplikacja łączy się z ECU pojazdu przez interfejs OBD-II, używając Bluetooth lub Wi-Fi, a następnie rejestruje kluczowe metryki zachowania

kierowcy. Dane te są dostępne zarówno w czasie rzeczywistym, jak i po zakończeniu podróży. Autorzy monitorowali takie parametry pojazdu jak prędkość pojazdu, prędkość obrotowa i temperatura silnika, masowy przepływ powietrza i położenie przepustnicy.

Podsumowując, kierowca swoim zachowaniem na drodze może znacząco przyczynić się do redukcji śladu węglowego związanego z transportem drogowym. Kluczowymi aspektami w tym obszarze są prędkość pojazdu oraz dynamika jazdy. Nadmierna prędkość (agresywny styl jazdy) wiąże się najczęściej z większym zużyciem paliwa, co prowadzi do zwiększenia emisji CO₂. Jazda z umiarkowaną prędkością dostosowaną do warunków jazdy (zrównoważony styl), zazwyczaj w zakresie średniej wartości prędkości obrotowej silnika (zalecany punkt pracy silnika), jest bardziej efektywna pod względem zużycia paliwa. W kontekście dynamiki jazdy, gwałtowne przyspieszanie i hamowanie mogą znacznie zwiększyć zużycie paliwa [100], [101]. Z kolei łagodne i płynne przyspieszanie i stopniowe hamowanie pozwalają zredukować emisję CO₂.

2.4. Zużycie paliwa a emisja CO₂

Tematyka zużycia paliwa w kontekście modelowania emisyjności pojazdów samochodowych jest drugim wymienionym na rysunku 20 obszarem tematycznym. W przeciwieństwie do szeroko występujących problemów klasyfikacji/identyfikacji, dotyczących modelowania stylu jazdy kierowców, w przypadku zużycia paliwa dominują metody regresyjne, pozwalające na prognozowanie zużycia paliwa w pojazdach samochodowych. W literaturze odnaleźć można różne modele prognozowania zużycia paliwa na podstawie wybranych czynników i można je podzielić w następujący sposób [124]:

- oparte na mapowaniu silnika – dokładne, ale wymagające szerokiego zakresu danych dla każdego samochodu;
- oparte na regresji – opracowane na podstawie wyników regresji z przeprowadzonych eksperymentów, wykazują ograniczoną interpretację fizyczną;
- modele fizyczne – oparte na równaniach fizycznych opisujących zjawiska wpływające na zużycie paliwa, są one wymagające obliczeniowo;
- modele hybrydowe – łączą cechy ww. modeli w celu optymalizacji czasu i dokładności obliczeń (należą do najczęściej stosowanych).

Dostępne publikacje można podzielić także ze względu na rodzaj zastosowanej metody prognozowania zużycia paliwa, co zostało przedstawione w tabeli 8.

Tabela 8. Metody prognozowania zużycia paliwa w pojazdach samochodowych

Metoda prognozowania	Źródła
Regresja liniowa	[12], [44], [98]
Lasy losowe	[1]
Sieci neuronowe	[1], [44], [87], [98], [127], [129]
Głębokie uczenie	[44]
Drzewa decyzyjne	[1], [98], [123]
Logika rozmyta	[114]
Uczenie maszynowe	[2], [44], [127], [87]
Metody statystyczne	[44], [127]

Szeroko stosowane są metody regresji liniowej, modele uczenia maszynowego, sieci neuronowe, modele drzew decyzyjnych, metody głębokiego uczenia, logika rozmyta, czy metody statystyczne (tabela 8). W dalszej części niniejszego podrozdziału szczegółowo zostanie omówione zastosowanie wymienionych metod w kontekście zużycia paliwa, prezentowane w literaturze.

W artykule [44] przeprowadzono badanie predykcyjne dotyczące zużycia paliwa i emisji CO₂. Wykorzystując uczenie maszynowe, głębokie uczenie, metody statystyczne, przedstawiono kompleksowy obraz zmienności zużycia paliwa ze względu na parametry techniczne. Techniki uczenia maszynowego i głębokiego uczenia umożliwiły również określenie korelacji między specyfikacjami pojazdu, a całkowitym zużyciem paliwa.

W artykule [98] wykazano, iż predykcja chwilowego zużycia paliwa jest możliwa przy pomocy różnych modeli, tj. regresji wielorakiej, drzew decyzyjnych czy sieci neuronowych bazujących jedynie na kilku wybranych zmiennych wejściowych. Autorzy artykułu [98] opracowali model obliczania chwilowego zużycia paliwa na podstawie dostępnych danych z portu OBD-II pojazdu. Celem analizy była weryfikacja wybranych modeli przy uwzględnieniu wybranych czynników, tj. obciążenia silnika, prędkości kątowej wału korbowego silnika, prędkości pojazdu, nachylenia terenu, przyspieszenia pojazdu oraz określenia odchylenia pomiędzy wartością przewidywaną podaną przez model a wartością zaobserwowaną.

Z kolei w artykule [127] przeprowadzono badanie dotyczące metod predykcji zużycia paliwa przez pojazdy za pomocą algorytmów i modeli statystycznych z uwzględnieniem sieci neuronowych i tradycyjnego uczenia maszynowego. Z przeprowadzonych badań wynika, że model zużycia paliwa oparty na sieciach neuronowych charakteryzuje się dużą dokładnością

tylko wtedy, gdy dostępne są wystarczające duże zbiory danych wejściowych. Ponadto zastosowanie hybrydowych modeli zużycia paliwa, składających się z wielu metod uczenia maszynowego, umożliwia uzyskanie najbardziej wiarygodnych wyników spośród analizowanych metod.

W artykule [129] wykorzystano sztuczne sieci neuronowe (ANN) do przewidywania zużycia paliwa w samochodach osobowych na podstawie ich parametrów technicznych. Autorzy stworzyli model predykcyjny ANN, analizując dane z ok. 2 tysięcy pojazdów wyprodukowanych w ostatniej dekadzie. Skupili się na określeniu wpływu na zużycie paliwa różnych zmiennych, takich jak m.in. objętości skokowej silnika, liczby cylindrów czy masy pojazdu. Wykorzystano różne konfiguracje sieci neuronowych, a skuteczność modeli oceniono na podstawie współczynnika korelacji Pearsona i współczynnika determinacji R^2 .

Autorzy artykułu [87] wykorzystali metody uczenia maszynowego, w tym sieci neuronowe LSTM (z ang. Long Short-Term Memory), do analizy i przewidywania zużycia paliwa w oparciu o dane dotyczące stylu jazdy. Badanie przeprowadzono w oparciu o bazę danych uzyskaną z rzeczywistych tras z uwzględnianiem informacji o warunkach drogowych. Wyniki pokazują, że określone zachowania kierowców mają znaczący wpływ na zużycie paliwa, a wykorzystane metody mogą skutecznie klasyfikować i przewidywać krótkoterminowe zużycie paliwa. Wyniki modelu LSTM zostały porównane z wynikami innych metod uczenia maszynowego, co pozwoliło na ocenę jego skuteczności i dokładności w kontekście innych dostępnych technik. Wspomniane informacje wskazują na kompleksowy i wielowymiarowy proces analizy danych, który uwzględniał zarówno czasową dynamikę zachowań kierowców jak i porównanie różnych metod modelowania w celu poprawy efektywności przewidywań zużycia paliwa.

Algorytmy uczenia maszynowego, tj. modele maszyn wektorów nośnych (SVM) są również stosowane do estymacji parametrów opisujących zużycie paliwa w oparciu o dane z systemu diagnostycznego pojazdu OBD-II. W artykule [2] wykorzystano kilka pojazdów testowych na trasie pomiędzy dwoma miastami w Jordanii, badając związek między zużyciem paliwa a prędkością obrotową silnika oraz położeniem przepustnicy.

Wyniki artykułu [1] wskazują z kolei, iż model zużycia paliwa określony na podstawie danych z diagnostyki pokładowej (OBD-II), bazujący na algorytmie lasów losowych może zapewniać bardziej dokładną predykcję zużycia w porównaniu z modelami opartymi o sieci neuronowe.

W artykule [12] opracowano liniowy model regresji, który został wykorzystany do szacowania zużycia paliwa, wykorzystując dane z portu diagnostycznego pojazdu i jednostki

inercyjnej IMU. W badaniu wzięło udział ponad 100 kierowców, którzy łącznie przejechali ok. 8000 km. Omawiany liniowy model regresji wielorakiej wykorzystuje dostępne zmienne, tj. prędkość i przyspieszenie pojazdu oraz pozycję pedału gazu, do modelowania zużycia paliwa i jest szczególnie efektywny przy agregacji danych, dlatego może znaleźć zastosowanie w systemach wspomagania kierowcy oraz symulacjach środowiskowych.

Odmienne podejście zaprezentowano w artykule [44], gdzie wykorzystano dane dotyczące ok. 5 tysięcy pojazdów samochodowych wyprodukowanych w latach 2017–2021. Skupiono się na porównaniu marek i modeli samochodów pod kątem zużycia paliwa i emisji CO₂. Wykorzystano regresję wielomianową, liniową, wielokrotną, logarytmiczną i algorytmy głębokiego uczenia, w tym sieci konwolucyjne (CNN), do przewidywania zużycia paliwa i emisji CO₂ dla przyszłych projektów pojazdów. Wyniki pokazują dokładność tych metod wynoszącą do 98,6% dla regresji wielomianowej i około 75% dla innych modeli.

Z kolei w artykule [95] autorzy skupili się na analizie statystycznej dotyczącej zużycia paliwa z wykorzystaniem danych z interfejsu OBD-II dla pojazdu hybrydowego. Badanie obejmowało eksperymenty z rzeczywistych cykli jazdy z wykorzystaniem danych z układu GNSS oraz OBD-II w przypadku hybrydowego pojazdu samochodowego. Celem badania była ocena dynamiki jazdy kierowcy oraz bezpieczeństwa ruchu drogowego. Zastosowano analizę statystyczną, skupiając się na miarach prędkości, przyspieszenia i średniej siły ciągłej (MTF), aby ustalić, jak styl jazdy wpływa na zużycie paliwa i emisję CO₂.

Natomiast w artykule [67] przedstawiono w jaki sposób dane uzyskane z systemu diagnostycznego pojazdu mogą służyć do oceny stylu jazdy kierowcy pod kątem zużycia paliwa. W badaniu dokonano analizy przebiegów m.in. prędkości pojazdu, prędkości obrotowej silnika, pozycji pedału przyspieszenia, pozycji przepustnicy oraz momentu obrotowego silnika. Testy przeprowadzono w realnych warunkach drogowych, a wyniki pozwoliły na ocenę różnic w stylach jazdy kierowców w warunkach miejskich i pozamiejskich. Badanie potwierdziło przydatność danych z OBD-II w ocenie stylu jazdy i jego wpływu na zużycie paliwa i emisję zanieczyszczeń.

Podsumowując, interfejs OBD jest powszechnie stosowany do ścisłego monitorowania wybranych parametrów pracy silnika oraz innych, kluczowych charakterystyk pracy pojazdu, co umożliwia jego zastosowanie także w celu modelowania zużycia paliwa. Analiza danych z interfejsu pozwala na bardziej świadomą i ekonomiczną eksploatację pojazdów, a także może przyczyniać się do ochrony środowiska przez minimalizację zużycia paliwa w pojazdach samochodowych.

Zauważyć należy, że do matematycznego modelowania problemu zużycia paliwa wykorzystuje się szereg różnych metod. Są to zarówno złożone algorytmy uczenia maszynowego jak również prostsze rozwiązania, jak np. regresja liniowa. Niemniej jednak większość badań koncentruje się na eksperymentach przeprowadzanych w kontrolowanych warunkach, natomiast rozwiązań dedykowanych pojedynczemu kierowcy nie jest wiele. Tymczasem skuteczne wdrażanie zmian na poziomie globalnym wymaga zaangażowania każdego potencjalnego użytkownika ruchu drogowego. W związku z tym, opracowanie narzędzi dedykowanych indywidualnym użytkownikom stało się główną motywacją niniejszej pracy.

2.5. Badanie systemów diagnostycznych w pojazdach samochodowych

Doskonalenie procesu zużycia paliwa i jego ciągłe monitorowanie jest przedmiotem wielu badań naukowych. Poruszające się po drogach na całym świecie, mocno wyeksploatowane pojazdy z silnikami spalinowymi, które nie spełniają obecnie obowiązujących standardów ekologicznych, mogą stanowić poważne zagrożenie. Jednym ze skutecznych sposobów zapobiegania gwałtownemu wzrostowi emisji z takich pojazdów jest wczesne wykrywanie usterek układu zasilania i sterowania silnikiem, co stanowiło genezę systemów pokładowej diagnostyki silnika [68]. Monitorowanie emisji CO₂ oraz zużycia paliwa umożliwiają pokładowe systemy diagnostyczne. W dalszej części podrozdziału tematyka dotycząca możliwego użycia systemów diagnostycznych do monitorowania emisji CO₂ zostanie poddana szczegółowej analizie.

Pojęcie systemów diagnostycznych OBD (z ang. On-Board Diagnostics) jest zagadnieniem złożonym, ponieważ systemy te stanowią zbiór różnych testów diagnostycznych oraz procedur obliczeniowych i decyzyjnych, które są wykonywane w pojeździe w czasie rzeczywistym. Głównym ich celem jest ocena sprawności emisyjnej oraz układów odpowiedzialnych za bezpieczeństwo czynne (układ przeciwblokujący – ABS), układ przeciwpółślizgowy – ASR, układ stabilizacji toru jazdy – ESP itd.) i bierne pojazdu (poduszki powietrzne, pasy bezpieczeństwa, konstrukcja stref zgniotu itd.). Ważną kwestią dotyczącą pojęcia systemów diagnostycznych jest problem ich nazewnictwa, mogący prowadzić do nieporozumień, ponieważ wspomniany system występuje w USA pod nazwą OBD-II, natomiast w Europie nosi nazwę EOBD (z ang. Euro-On-Board-Diagnostic) [78].

Wymienione OBD-II jest terminem używanym w kontekście prawnym oraz technicznym w celach określenia zbioru przepisów oraz charakterystyki systemów

pokładowych wykonanych zgodnie z wymaganiami normy. Analogicznie jest w przypadku europejskiego standardu systemów diagnostycznych EOBD, który stanowi rozwinięcie standardu OBD-II. Wynika to z faktu, iż system OBD-II (jako następca OBD-I) zaczął obowiązywać w USA dla wszystkich samochodów z silnikami benzynowymi od 1996 roku, a od 1997 roku z silnikami o zapłonie samoczynnym. W Europie natomiast w tym czasie nie było żadnych podobnych uregulowań i rozwiązań. System OBD-II, bardzo szybko po wprowadzeniu, stał się światowym standardem. Natomiast w Europie na jego bazie utworzono wspomniany EOBD, który zaczął obowiązywać od 2000 roku dla samochodów z silnikami spalinowymi i od 2004 roku z silnikami o zapłonie samoczynnym [78].

Analizując tematykę związaną z systemami pokładowymi OBD nie można pominąć niezwykle ważnego pojęcia jakim jest diagnostyka pojazdów, którą można zdefiniować jako proces lokalizacji elementów lub układów, które z różnych przyczyn (zużycie, uszkodzenie itp.) nie mogą spełniać swojej funkcji zgodnej ze specyfikacją producenta. Omawiana diagnostyka może odbywać się m.in. w stacjach diagnostyczno-naprawczych lub stacjach kontroli pojazdów. Zazwyczaj wynika ona z zauważonych przez użytkownika uszkodzeń lub może być, np. efektem obowiązkowej okresowej kontroli pojazdów [78]. Uszkodzenia pojazdu można z kolei podzielić ze względu na wywołane skutki w następujący sposób [78]:

- uszkodzenia emisyjne – skutkują zwiększoną emisją związków toksycznych z układu wylotowego lub zwiększonym zużyciem paliwa;
- uszkodzenia wpływające na bezpieczeństwo jazdy – dotyczą najczęściej elementów podwozia oraz nadwozia;
- uszkodzenia nieemisyjne układu napędowego – skutkują pogorszeniem dynamiki pojazdu;
- uszkodzenia wpływające na komfort jazdy – dotyczą najczęściej elementów i układów nadwozia.

Kierowcy najszybciej zauważają i likwidują uszkodzenia nieemisyjne układu napędowego z uwagi na zauważalne pogorszenie dynamiki pojazdu. Również uszkodzenia wpływające na komfort jazdy są odczuwalne dla kierowcy, w przeciwieństwie do uszkodzeń emisyjnych oraz części uszkodzeń związanych z bezpieczeństwem, które najczęściej nie wykazują żadnych zauważalnych symptomów (tylko w sytuacjach krytycznych). Specyfika uszkodzeń emisyjnych sprawia, iż emisja związków toksycznych z pojazdów samochodowych może być w wielu przypadkach niekontrolowana. Pomimo starań producentów pojazdów podczas fazy projektowania i produkcji samochodów, emisja może zwiększyć się w czasie eksploatacji. Mechanizmem, który wykrywa uszkodzenia już we wczesnej fazie ich

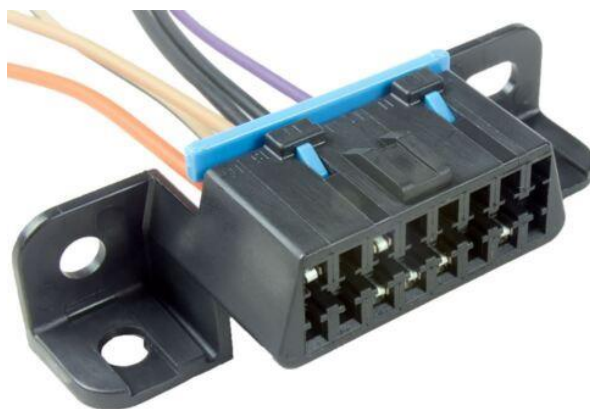
występowania jest właśnie system diagnostyki pokładowej, stanowiący podmiot omawianej normy OBD-II/EOBD [78].

Norma ta zakłada, że każdy system OBD powinien spełniać trzy podstawowe cechy, które są kluczowe dla realizacji funkcjonalności systemu [78]:

- 1) Uniwersalność – możliwość rozbudowy o nowe elementy i podsystemy wraz z możliwością przeniesienia na inne typy pojazdów.
- 2) Efektywność – możliwość wykrycia i identyfikacji uszkodzeń.
- 3) Dostępność – możliwość organizacji i pracy według przyjętych standardów.

System OBD, według przyjętych wymagań formalnych, umożliwia łatwe rozszerzanie systemu na nowe podsystemy i elementy, co zapewnia uniwersalność i stanowi ogromną zaletę rozwiązania (cecha nr 1). Druga z wymienionych cech, tzn. efektywność, zapewnia możliwość diagnozowania i lokalizacji uszkodzeń wszystkich elementów wpływających na emisję. W omawianym kontekście elementem niesprawnym jest zatem taki element, którego praca może wywołać znaczny wzrost emisji związków toksycznych. Z kolei dostępność (cecha nr 3) pozwala użytkownikom uzyskać dostęp do informacji diagnostycznej oraz serwisowej. Pierwsza z wymienionych informacji, przez dane z systemu pomiarowego, pozwala na lokalizację uszkodzonych elementów, natomiast druga z wymienionych umożliwia dostęp do danych producenta dotyczących możliwości naprawy uszkodzenia lub wymiany uszkodzonego elementu [78]. Podsumowując, na podstawie przytoczonych informacji, można zauważyć ogromne możliwości systemu diagnostyki OBD zarówno pod kątem wykrycia, identyfikacji, lokalizacji, ale również naprawy uszkodzonego elementu.

Praktycznym aspektem dotyczącym pokładowych systemów diagnostycznych są czytniki informacji diagnostycznej, stanowiące warstwę sprzętową omawianych systemów. Omawiane normy OBD definiują zasady ich komunikacji z siecią samochodu przez 16-stykowe złącze diagnostyczne DLC (z ang. Data Link Connector, rys. 21). Wspomniany czytnik pozwala na monitorowanie informacji dotyczących, m.in. parametrów silnika, pojawiających się błędów wyrażonych przez odpowiednie kody itd. W konsekwencji przyjętych norm, złącze DLC powinno być umieszczone w kabinie pojazdu w odległości 0,5 m od osi kierownicy, w łatwo dostępnym miejscu, cechującym się dobrą widocznością [78].



Rys. 21. Widok złącza DLC [106]

System OBD generuje informację diagnostyczną, która jest przekazywana kierowcy za pomocą kontrolki MIL (z ang. Malfunction Indicator Light) na tablicy rozdzielczej samochodu. Jest to podstawowy sposób komunikacji z kierowcą pojazdu, informujący o występujących problemach. Użytkownicy natomiast mogą uzyskać dostęp do informacji diagnostycznej pojazdu przez podłączenie odpowiedniego czytnika OBD do złącza DLC w samochodzie. W rozpatrywanym kontekście wymienione czytniki stanowią urządzenia spełniające zespół zdefiniowanych w normach wymagań funkcjonalnych [78].

Normy wskazują również dodatkowe pojęcie bezpośrednio związane z czytnikami OBD, tzw. monitor diagnostyczny, który oznacza procedurę diagnostyczną centralnego komputera sterującego, wykonywaną w celu weryfikacji poprawności pracy elementu lub funkcji podsystemu samochodu. Monitory wykorzystują testy do oceny stanu technicznego pojazdu. Zadaniem monitora jest również przechowywanie wyników wspomnianych testów i ewentualnego powiadamiania o wystąpieniu uszkodzenia wpływającego na emisję z układu wylotowego lub zasilania. Istotne jest również to, iż każdy monitor dotyczy wyłącznie jednego elementu lub podsystemu wpływającego na emisję co powoduje, iż liczba występujących w danym samochodzie monitorów może się od siebie różnić w zależności od rodzaju silnika i zastosowanych technologii systemu kontroli emisji. Monitory można podzielić ze względu na ciągłość ich pracy na ciągłe i warunkowe (nieciągłe). Monitory ciągłe dotyczą elementów i podzespołów, które mogą być rejestrowane na bieżąco, bez wpływu na działanie pozostałych monitorów (np. spalanie paliwa w silniku). Monitory warunkowe dotyczą elementów wymagających znacznie dłuższego czasu obserwacji, np. praca reaktora katalitycznego [130].

Wspomniane testy i monitory są wykonywane niezależnie od podpiętego czytnika OBD. Zakres dostępnych dla użytkownika informacji definiuje norma OBD-II/EOBD, która wprowadza 9 podstawowych trybów testowania [24]:

- Tryb \$1 – dostęp do bieżących wartości parametrów układu napędowego oraz informacji systemowych.
- Tryb \$2 – dostęp do informacji tzw. „zamrożonej ramki”, danych przechowywanych w pamięci sterownika, dotyczących informacji określającej stan pojazdu w chwili wystąpienia uszkodzenia.
- Tryb \$3 – dostęp do kodów błędów zarejestrowanych przez sterowniki pojazdu.
- Tryb \$4 – możliwość kasowania wszystkich kodów błędów zapisanych w pamięci sterownika.
- Tryb \$5 – testowanie czujników tlenu.
- Tryb \$6 – testowanie monitorów warunkowych.
- Tryb \$7 – testowanie monitorów ciągłych.
- Tryb \$8 – kontrola nad systemem diagnostyki.
- Tryb \$9 – odczyt informacji o pojeździe [24].

Tryb \$1 jest najważniejszym trybem systemu, ponieważ umożliwia dostęp do chwilowych wartości podstawowych parametrów pojazdu, m. in. prędkość obrotowa silnika, prędkość pojazdu, obciążenie silnika, temperatura płynu chłodzącego, krótkoterminowa oraz długoterminowa korekta mieszanki paliwowej, ciśnienie w kolektorze dolotowym, pozycja przepustnicy, temperatura powietrza dolotowego, masowy przepływ powietrza, napięcie sondy lambda, ciśnienie paliwa itd. Wartości parametrów możliwych do odczytu w trybie \$1 posiadają częstotliwość próbkowania nie mniejszą niż 2 Hz, natomiast dane dotyczące tzw. „ramki zamrożonej” (tryb \$2) przedstawiają jedynie wartości wybranych parametrów trybu \$1, zarejestrowanych w chwili wystąpienia uszkodzenia. Z kolei tryby \$3 i \$4 dotyczą kodów błędów i umożliwiają kolejno odczyt i kasowanie wspomnianych błędów. Kolejne tryby \$5, \$6, \$7 przedstawiają wyniki testów monitora czujników tlenu (np. sonda lambda, tryb opcjonalny), monitorów warunkowych (np. monitor reaktora katalitycznego, układu odprowadzania par paliwa itd.) i ciągłych (np. wartości elementów emisyjnych pojazdu). Tryb \$8 stanowi wsparcie dla diagnosty po skasowaniu błędów, natomiast tryb \$9 umożliwia, m.in. odczyt numeru identyfikacyjnego VIN pojazdu [78]. Oprócz omówionych 9 podstawowych trybów pracy \$1–\$9 systemu OBD występuje także szereg rozszerzonych trybów pracy (\$10–\$33) umożliwiających dostęp do zaawansowanych informacji dotyczących stanu pozostałych podsystemów pojazdu [78].

Wspomniane tryby zapewniają dostęp do szeregu informacji przez stosowny czytnik systemów diagnostycznych, który występuje w różnych odmianach. Na podstawie przeprowadzonej analizy rynkowej takich urządzeń można wyróżnić, m. in.:

- czytniki kodów OBD
- skanery uniwersalne OBD
- profesjonalne skanery OBD,
- interfejsy bezprzewodowe OBD.

Pierwsze wymienione urządzenia, tzn. czytniki kodów OBD, stanowią niewielkie i nieskomplikowane urządzenia, używane głównie do odczytu i kasowania podstawowych kodów błędów diagnostycznych. Składają się najczęściej z niewielkiego wyświetlacza wraz z dołączoną klawiaturą oraz przewodem zakończonym wtykiem OBD (rys. 22).



Rys. 22. Czytnik kodów OBD Technaxx TX-184 [107]

Urządzenie posiada ograniczoną funkcjonalność i zazwyczaj dostarcza jedynie ogólnych informacji o kodzie błędu, bez szczegółowej diagnostyki. Nie umożliwia przy tym dostępu do informacji dotyczących monitorowania emisji CO₂.

Bardziej złożonymi urządzeniami są uniwersalne skanery OBD. Oprócz podstawowego odczytu i kasowania błędów umożliwiają one także m.in. diagnostykę elementów pojazdu, dostęp do wybranych parametrów w czasie rzeczywistym oraz wykonanie testów wybranych układów pojazdu. Skaner uniwersalny OBD (rys. 23) składa się najczęściej z kolorowego wyświetlacza, dołączonej klawiatury oraz przewodu zakończonego wtykiem OBD. Urządzenie posiada dodatkowe wyprowadzenia w postaci przewodu USB, które umożliwia komunikację z komputerami PC za pomocą dedykowanego oprogramowania.



Rys. 23. Skaner uniwersalny OBD 24DIAG S403 [109]

Z uwagi na to, iż urządzenie oferuje dodatkową funkcjonalność oraz większe możliwości diagnostyczne w porównaniu do czytników błędów OBD, to jest ono dedykowane mechanikom samochodowym, przeprowadzającym stacjonarne testy diagnostyczne. Skaner umożliwia dostęp do informacji dotyczących monitorowania emisji CO₂. Urządzenie jednak nadal może nie oferować niektórych wybranych, zaawansowanych funkcji diagnostycznych, dostępnych jedynie dla skanerów profesjonalnych, przeznaczonych konkretnym producentom pojazdów.

Przykładem wymienionego skanera profesjonalnego jest, np. skaner dedykowany pojazdowi marki Mazda oraz Ford (rys. 24).



Rys. 24. Skaner OBD Launch Creader Elite dedykowany do samochodów marek Ford, Mazda [110]

Wspomniane skanery, oprócz podstawowej wymienionej funkcjonalności uniwersalnych skanerów OBD- umożliwiają obsługę większości modułów sterujących, w jakie wyposażony jest określony pojazd, np. ABS, ESP, ASR, TPMS, klimatyzacja, tempomat itd. Ponadto zapewniają kompleksowe funkcje diagnostyczne i dlatego znalazły zastosowanie

w profesjonalnych warsztatach samochodowych. Z uwagi na rozbudowaną funkcjonalność wymagają jednak znacznie większej wiedzy technicznej w celu skutecznej pracy, a ich niewłaściwa obsługa grozi uszkodzeniem pojazdu.

Kolejnym, obecnie niezwykle popularnym rozwiązaniem, które zapewnia pełną mobilność (brak przewodów), uniwersalną funkcjonalność, prostotę działania oraz niewielkie rozmiary są bezprzewodowe interfejsy OBD (rys. 25).



Rys. 25. Interfejs bezprzewodowy Vgate iCar PRO WiFi [108]

Interfejsy najczęściej bazują na mikrokontrolerze opartym o układ ELM327 (nie posiadają wyświetlacza) i umożliwiają bezprzewodowe połączenie (Wi-Fi, Bluetooth) ze smartfonem lub tabletem. Są kompatybilne z różnymi aplikacjami (Android, iOS), oferującymi zróżnicowaną funkcjonalność. Stanowią wygodne rozwiązanie - w porównaniu z najczęściej stosowanymi w stacjonarnych testach, opisanymi wcześniej, uniwersalnymi oraz profesjonalnymi skanerami OBD. Interfejsy bezprzewodowe umożliwiają dostęp do informacji dotyczących monitorowania emisji CO₂.

Podsumowując, skanery OBD stanowią ważne narzędzie w diagnostyce pojazdów, ponieważ pozwalają na wykrywanie i lokalizowanie różnych uszkodzeń w samochodach, w tym emisyjnych. Dzięki uniwersalności systemu OBD, skanery są kompatybilne z większością współczesnych pojazdów samochodowych, niezależnie od marki, modelu, czy typu silnika. Efektywność systemu OBD umożliwia szybkie identyfikowanie uszkodzeń, co znacząco skraca czas potrzebny na naprawę i minimalizuje ryzyko poważniejszych awarii. Dostępność systemu dla przeciętnych użytkowników jest kolejną zaletą systemu.

Jedną z najważniejszych cech skanerów OBD jest ich zdolność do monitorowania i testowania różnych podzespołów pojazdu, w tym elementów układu napędowego, układów sterujących silnika, reaktora katalitycznego czy sondy lambda. Skanery OBD umożliwiają ich testowanie oraz szczegółową analizę danych diagnostycznych. Tryby te obejmują między

innymi odczyt kodów pojawiających się błędów, monitorowanie parametrów w czasie rzeczywistym czy też testy emisji spalin.

Interfejsy bezprzewodowe OBD są niezwykle szybko rozwijającą się technologią, umożliwiającą dostęp do pokładowego systemu diagnostyki pojazdu. W przeciwieństwie do przewodowych skanerów OBD pozwalają na bezprzewodowy dostęp do danych diagnostycznych pojazdu, co ułatwia monitorowanie jego stanu technicznego w czasie rzeczywistym i ocenę emisji CO₂, w tym również w formie zdalnej (na odległość). Ponadto interfejsy te mogą być stosowane przy pomocy smartfonów do rejestracji i archiwizacji przebiegów parametrów pojazdu w czasie rzeczywistym, w trakcie jazdy, w celu ich późniejszego badania i oceny stanu technicznego pojazdu.

3. CEL PRACY I ZAŁOŻENIA BADAWCZE

Przeprowadzona kwerenda literatury wykazała, że większość badań w obszarze kształtowania poziomu emisji CO₂ dotyczy rozwiązań na poziomie globalnym, tymczasem brak jest narzędzi wspierających zaangażowanie pojedynczych użytkowników pojazdów osobowych w tym zakresie. W pracy skoncentrowano się na zagadnieniach związanych z pojazdami wyposażonymi w silniki o zapłonie iskrowym (ZI).

Przegląd istniejącej literatury oraz doświadczenie autora niniejszego opracowania wykazały istnienie luki badawczej potwierdzającej niewielką liczbę metod, technik i narzędzi dedykowanych kompleksowemu modelowaniu emisji CO₂ z pojazdów samochodowych poświęconych indywidualnym użytkownikom, co ogranicza możliwości dokładnego przewidywania emisji oraz identyfikacji kluczowych czynników wpływających na jej poziom. W szczególności brakuje zintegrowanych podejść, które łączyłyby różnorodne zmienne objaśniające, takie jak styl jazdy kierowcy, parametry techniczne pojazdu oraz warunki eksploatacyjne, z nowoczesnymi metodami analizy danych, co uzasadnia potrzebę dalszych badań w tym obszarze.

W kontekście rosnącej świadomości ekologicznej oraz konieczności redukcji emisji gazów cieplarnianych, istotnym problemem badawczym jest zrozumienie i modelowanie emisji dwutlenku węgla (CO₂) z pojazdów osobowych. W konsekwencji powyższych rozważań sformułowano następujący problem badawczy (pytanie badawcze):

W jaki sposób można wpływać na emisję dwutlenku węgla z pojazdów osobowych na poziomie pojedynczego użytkownika?

Dla jego odpowiedzi określono problemy częściowe:

- 1. Jakie mierzalne i sterowalne na poziomie użytkownika pojazdu osobowego czynniki mają największy wpływ na poziom emisji CO₂?*
- 2. W jaki sposób wbudowane systemy diagnostyczne w pojeździe mogą wspierać monitorowanie emisji CO₂ i jakie mają znaczenie dla modelowania parametrów eksploatacyjnych pojazdów?*
- 3. Jakie praktyczne rekomendacje wynikają z uzyskanych modeli dla pojedynczego użytkownika ruchu oraz strategii obniżenia emisji gazów cieplarnianych w kontekście zrównoważonego rozwoju?*

Możliwość rozpoznania i oceny wpływu różnych czynników na emisję CO₂ z pojazdów osobowych oraz wykorzystania zastosowanych metod modelowania stały się fundamentem do

postawienia celu głównego rozprawy, którym jest **opracowanie modelu oceny kształtowania poziomu emisji CO₂ na podstawie informacji pozyskanych z interfejsów diagnostycznych.**

Analiza zgromadzonych danych, dokonanie oceny stanu rzeczywistego oraz zbudowanie modelu emisji CO₂ z pojazdów osobowych pozwoli na identyfikację kluczowych czynników wpływających na poziom emisji, a także na opracowanie skutecznych strategii ich redukcji na poziomie pojedynczego użytkownika pojazdu. Dzięki zastosowaniu odpowiednich analiz statystycznych oraz metod matematycznego modelowania, możliwe będzie nie tylko przewidywanie przyszłych wartości emisji, ale także formułowanie rekomendacji dla polityki transportowej oraz praktyk eksploatacyjnych, co przyczyni się do zrównoważonego rozwoju i ochrony środowiska. Powyższe stanowi podstawę do sformułowania następującej tezy badawczej:

Możliwe jest opracowanie modelu matematycznego kształtowania poziomu emisji CO₂ z wykorzystaniem dostępnych w pojeździe osobowym systemów dla pojedynczego użytkownika.

Dla jej zrealizowania określono także cele szczegółowe:

1. *Określenie mierzalnych i sterowalnych czynników mających wpływ na poziom emisji CO₂ na poziomie użytkownika pojazdu samochodowego oraz ich analiza.*
2. *Zbadanie roli systemów diagnostycznych w monitorowaniu i analizie emisji CO₂ oraz opracowanie modeli emisji CO₂ w odniesieniu do poszczególnych trybów jazdy z wykorzystaniem interfejsu OBD wraz z określeniem wpływu wybranych parametrów eksploatacyjnych pojazdu na wyznaczanie poziomu emisji CO₂.*
3. *Opracowanie rekomendacji dla pojedynczego użytkownika ruchu oraz wytycznych redukcji gazów cieplarnianych w odniesieniu do strategii zrównoważonego rozwoju.*

Realizacja powyższych celów pozwoli na uzyskanie wiedzy dotyczącej emisji CO₂ z pojazdów osobowych oraz pozwoli na dostarczenie narzędzi modelowania i przewidywania tych emisji, co jest kluczowe dla działań proekologicznych.

W rozprawie wykorzystano zarówno metody teoretyczne jak i empiryczne, ze względu na specyfikę i wymagania badanych zagadnień. Do głównych metod teoretycznych stosowanych w rozprawie należą:

- Analiza literatury przedmiotu – w ramach której dokonano przeglądu i analizy dotychczasowych publikacji naukowych, co pozwoliło na ocenę stanu wiedzy w danym obszarze, identyfikację luk badawczych oraz odnalezienie inspiracji do formułowania nowych pytań i tez.

- Synteza – która umożliwiła łączenie różnych teorii i koncepcji w spójny model teoretyczny lub wspomaga wyciąganie ogólnych wniosków na podstawie zróżnicowanych źródeł i perspektyw badawczych.
- Dedukcja i indukcja – które zostały wykorzystane do wyprowadzania szczegółowych wniosków z ogólnych teorii i tworzenia ogólnych zasad na podstawie obserwacji i specyficznych przypadków.
- Modelowanie teoretyczne – opracowanie modeli matematycznych, które pozwalają na testowanie przyjętych założeń badawczych i sprawdzanie relacji między zmiennymi. W badaniach wykorzystano wybrane techniki statystyczne, np. regresja wieloraka, czy regresyjne drzewa decyzyjne, pozwalające na identyfikację zależności między zmiennymi oraz ocenę wpływu różnych czynników na emisję CO₂.

Do głównych metod empirycznych zastosowanych w rozprawie należały badania eksperymentalne. Otrzymane w ich wyniku dane dotyczyły wybranych parametrów ruchu pojazdu uzyskanych z systemów diagnostycznych OBD (On-Board Diagnostics) oraz obserwacje zachowań kierowców, które były badane w ramach dysertacji.

Eksperymenty realizowano z zastosowaniem metody obserwacji, której wynikiem jest arkusz obserwacji zawierający rejestrowane pomiary.

Założenia badawcze

Prowadzone badania eksperymentalne zakładały długoterminową rejestrację wybranych parametrów poruszającego się pojazdu samochodowego. Wybrano w tym celu samochód z silnikiem o zapłonie iskrowym, którym kierowało sumarycznie sześciu zróżnicowanych pod względem wieku oraz doświadczenia kierowców. Taka liczba kierowców pozwoliła na uwzględnienie różnych stylów jazdy. Pomiary były prowadzone przez 3 lata, a obszarem badania był teren Polski. Całkowity dystans przejechany przez samochód wynosił ok. 12 tys. km. W każdym miesiącu prowadzonych badań (36 miesięcy) kierowcy, którzy zmieniali się losowo, mieli za zadanie przejechać min. 300 km. W celu ochrony prywatności poszczególnych kierowców, dane dotyczące ich przejazdów zanonimizowano. Większość tras była planowana z wyprzedzeniem (szczególnie na drogach ekspresowych oraz autostradach), dobierano je w taki sposób, aby były możliwie najbardziej zróżnicowane. Trasy zawierały liczne odcinki miejskie, drogi ekspresowe, ale również częste zatrzymania (wyłączenia/włączenia silnika), co umożliwiło zarejestrowanie różnych trybów jazdy pojazdu. Kierowcy otrzymali wytyczne, aby prowadzić pojazd w sposób odzwierciedlający ich codzienne nawyki jazdy oraz nie używać tempomatu w trakcie jazdy.

W celu akwizycji danych pomiarowych dotyczących wybranych parametrów technicznych pojazdu samochodowego wykorzystano interfejs diagnostyczny OBD. Warstwę sprzętową stanowiły bezprzewodowy adapter OBD wraz ze smartfonem, z kolei dedykowana aplikacja stanowiła niezbędne do komunikacji oprogramowanie. Koncepcja przyjętego schematu pomiarowego wykorzystanego w badaniach została przedstawiona na rysunku 26.



Rys. 26. Koncepcja schematu pomiarowy wykorzystywanego w badaniach

Pojazdem użytym do badań drogowych była Mazda 3 Sedan SkyActiv-G 2.0 (rys. 27), której dane techniczne podano w tabeli 9.



Rys. 27. Widok obiektu badań – Mazda 3 Sedan SkyActiv-G 2.0

Tabela 9. Dane techniczne wybranego pojazdu z silnikiem ZI [105]

Parametr	Dane
Producent	Mazda
Model	3
Rok produkcji	2016
Typ nadwozia	Sedan
Masa własna pojazdu	1280 kg
Rodzaj napędu	Na przednią oś
Paliwo	Benzyna
Pojemność zbiornika paliwa	51 dm ³
Objętość skokowa silnika	1998 cm ³
Liczba cylindrów	4
Układ cylindrów	Rzędowy
Skrzynia biegów	Manualna, 6 biegów do przodu, 1 bieg wsteczny
Maksymalna moc silnika	88 kW przy 6000 obr/min
Maksymalny moment obrotowy silnika	210 Nm przy 4000 obr/min
Opony	205/60 R16
Maksymalna prędkość	198 km/h
Przyspieszenie (od 0 do 100 km/h)	8,8 s
Emisja CO ₂	119 g/km
Norma emisji spalin	Euro 5

W celu rejestracji danych z pojazdu wykorzystano adapter Vgate iCar Pro OBD (rys. 28 a, b) wraz z aplikacją Torque Pro. Do rejestracji danych przez Wi-Fi wykorzystano telefon komórkowy LG G6 z systemem operacyjnym Android.



(a)

(b)

Rys. 28. Adapter OBD: (a) widok z przodu, (b) widok z tyłu

Kwerenda literatury wykazała, że na emisję CO₂ w samochodach spalinowych wpływa wiele czynników m.in. prędkość, przyspieszenie (opóźnienie) pojazdu, prędkość obrotowa silnika, moc silnika, moment obrotowy silnika, przełożenie skrzyni biegów, nachylenie terenu, obciążenie silnika, położenie przepustnicy itd., jednak ze względu na fakt, że genezą pracy było opracowanie narzędzia dedykowanego każdemu użytkownikowi pojazdu w badaniu wykorzystano jedynie czynniki bezpośrednio dostępne z interfejsu diagnostycznego OBD lub telefonu komórkowego oraz takie, na które bezpośredni wpływ ma styl jazdy kierowcy pojazdu. Z uwagi na powyższe w doborze zmiennych pominięto parametry techniczne opisujące, np. wymiary pojazdu, jego masę, pojemność skokową silnika oraz parametry, na które kierowca samochodu nie ma bezpośrednio wpływu, tj. temperatura powietrza dolotowego, temperatura płynu chłodzącego silnik itd. Odrzucono również parametry możliwe do obliczenia pośrednio (na podstawie innych parametrów), np. przyspieszenie pojazdu, aktualny bieg, nachylenie terenu itd. Skupiono się natomiast na wskaźnikach pracy silnika, tj. na prędkości obrotowej wału silnika, obciążeniu silnika, mocy silnika oraz kinematycznych parametrach pojazdu jak prędkość pojazdu.

Ostatecznie, na podstawie przyjętych kryteriów, w trakcie pomiarów rejestrowano bezpośrednio z interfejsu diagnostycznego wartości wymienionych wcześniej parametrów, tj.

- prędkość obrotową silnika,
- obciążenie silnika,
- moc silnika,
- prędkość pojazdu,
- moment obrotowy silnika,
- położenie przepustnicy,
- emisję CO₂.

Prędkość pojazdu rejestrowano przy pomocy wbudowanego w telefonie komórkowym odbiornika GNSS z uwagi na niską rozdzielczość pomiarów tego parametru (1 km/h) możliwych do odczytu z aplikacji Torque Pro. Aby zachować jakość surowych danych i uniknąć wpływu filtrowania na wyniki badań, zrezygnowano z procesu filtracji wspomnianych przebiegów. Ponadto z uwagi na to, że nie jest możliwe określenie poziomu emisji CO₂ podczas pracy w trakcie postoju pojazdu, nie uwzględniono tego typu obserwacji oraz pracy na biegu jałowym.

4. METODY MODELOWANIA MATEMATYCZNEGO ZASTOSOWANE W BADANIU

4.1. Teoretyczne aspekty modelowania

Na podstawie przeprowadzonej analizy stanu zagadnienia dotyczącej problematyki emisji CO₂ do atmosfery, można stwierdzić, iż poszukiwanie rozwiązań w zakresie ograniczenia negatywnego wpływu gazów cieplarnianych na środowisko jest kluczowym elementem prowadzonych obecnie badań. Autorzy publikacji naukowych postulują, że wzrost liczby pojazdów na świecie oraz zwiększona mobilność społeczeństwa przyczyniają się do pogłębiania zjawiska efektu cieplarnianego. Pomimo ciągłego rozwoju technologicznego, który powinien zmniejszać szkodliwość wspomnianego oddziaływania, problem pozostaje aktywny i ważny. Z kolei kwerenda literatury w zakresie stylu jazdy kierowców wskazuje, iż jest to jeden z głównych czynników wpływających na zużycie paliwa oraz emisję CO₂. Identyfikacja stylu jazdy kierowcy jest możliwa przez monitorowanie wartości wybranych parametrów pojazdu, tj. prędkość pojazdu, prędkość obrotowa silnika, obciążenie silnika itd. Możliwość odczytu wymienionych wartości zapewniają m.in. pokładowe systemy diagnostyczne OBD, stanowiące obowiązujący standard dla wszystkich nowych pojazdów spalinowych i elektrycznych. Umożliwiają one monitorowanie kinematycznych parametrów pojazdu, paramentów pracy silnika spalinowego oraz emisji CO₂, zapewniając jednocześnie rejestrację i archiwizację danych dotyczących pracy pojazdu. Dane te stanowią podstawę dalszych opracowań, umożliwiających modelowanie wybranych aspektów pracy pojazdów, np. emisji CO₂ w danych warunkach pracy silnika spalinowego.

Badanie i zrozumienie złożonych zjawisk przez ich uproszczone reprezentacje możliwe jest dzięki modelowaniu, które polega na opisie danego zjawiska lub właściwości. Ważnym aspektem jest możliwość modelowania określonego procesu zachodzącego w danym obiekcie, np. procesu zużycia paliwa i związanej z nim emisji gazów cieplarnianych. Badanie emisji CO₂ w pojazdach samochodowych stanowi przykład zastosowania koncepcji modelowania w kontekście obecnie występujących wyzwań środowiskowych [74], [119].

Modelowanie stanowi proces tworzenia reprezentacji zjawisk, obiektów lub procesów w warunkach rzeczywistych w formie zależności matematycznych, opisów koncepcyjnych lub wizualnych w celu m.in. ich zdefiniowania, zrozumienia, określenia ilościowego, wizualizacji czy symulacji. Według Encyklopedii Powszechnej PWN [33] modelowanie stanowi doświadczalną lub matematyczną metodę badania układów, zjawisk i procesów na podstawie konstruowania modeli. W przypadku modelowania doświadczalnego podstawą budowy modeli

jest podobieństwo fizyczne, natomiast w przypadku modelowania matematycznego proces polega na tworzeniu modeli matematycznych i zastosowaniu aparatu matematycznego w celu ich późniejszej analizy. Wyróżnia się trzy zasadnicze cele modelowania [70]:

- Prognostyczny – polega na wyznaczeniu wyjścia systemu przy znanym wejściu i funkcji transformacji.
- Identyfikacyjny – polega na utworzeniu opisów zasad (praw) funkcjonowania systemu, przy znanym wejściu i wyjściu.
- Racjonalizacyjny (np. optymalizacyjny) – umożliwia określenie warunków funkcjonowania systemu przy których charakterystyki ilościowe i jakościowe spełniają kryteria racjonalności, gdzie znane jest wyjście oraz funkcja transformacji, natomiast poszukiwane są wejścia.

Z modelowaniem związane są również dwa pojęcia, tj. oryginał i model. Oryginał stanowi rzeczywisty obiekt lub proces, który podlega badaniu, podczas gdy model jest jego uproszczoną reprezentacją. Według encyklopedii PWN [33] model oznacza teoretyczny (nominalny) lub fizyczny (realny) obiekt, który umożliwia określenie cech badanego obiektu, procesu lub zjawiska przez jego analizę lub obserwację. Z kolei związek modelu z procesem modelowania polega na tym, iż stanowi on końcowy efekt wymienionego procesu. Model stanowi metodę formalnego przedstawienia rozpatrywanej teorii, opisu pewnej doświadczalnej (empirycznej) obserwacji lub jest najczęściej połączeniem obu wymienionych aspektów. Najważniejszą cechą modelu jest jego zdolność do zastępowania badanego obiektu w procesie badań. Umożliwia to przeniesienie wyników uzyskanych doświadczalnie na rzeczywisty obiekt badań.

Koncepcja modelu jest ogólnym pojęciem podporządkowanym przyjętym przez badaczy celom. W literaturze można również odnaleźć inne definicje tego pojęcia, np. w pracy [14] przyjęto, iż model to taki układ, który odzwierciedlając lub odtwarzając przedmiot badań, zdolny jest zastępować go tak, że jego badanie dostarcza nowej informacji o tym przedmiocie. Z kolei autor pracy [118] przyjął, że model stanowi złożony obiekt, który umożliwia ustalenie przydatności oraz zastosowania dla odwzorowywanego fragmentu rzeczywistości. Podobna definicja została zaprezentowana w pracy [49], w której autorzy postrzegają model, jako substytut oryginału, przyjętą formę reprezentacji, wykorzystywaną do wyjaśnienia i przewidywania zachowania się rzeczywistego systemu, przy zachowaniu konkretnego celu badań.

Podział modeli zaproponowany w [48] wyróżnia następujące ich rodzaje, tj. fizyczne, matematyczne, schematyczne. Pierwszą wymienioną grupą modeli są modele fizyczne, które

charakteryzuje podobieństwo fizyczne pomiędzy oryginałem, a rozpatrywanym modelem. Wyróżnia je identyczna struktura fizyczna i funkcjonalna występująca zarówno w przypadku badanego obiektu jak również badanego zjawiska fizycznego, która jest łatwiejsza do obserwacji i sterowania. Główną zaletą modeli fizycznych jest ich użyteczność wynikająca ze zmiany skali, np. złożone struktury mogą być zmniejszane do rozmiarów umożliwiających ich późniejsze badanie.

Kolejną grupą modeli są modele matematyczne, które opisują dane zjawisko, proces lub obiekt za pomocą równania lub układów równań. Można je określić jako zbiór reguł i zależności, które umożliwiają na drodze obliczeń odzwierciedlenie procesów zachodzących w badanym obiekcie. Można je podzielić na analityczne oraz numeryczne. W przypadku modeli analitycznych rozwiązanie danego problemu jest możliwe przez użycie metod algebraicznych, różniczkowych lub całkowych, natomiast w przypadku modeli numerycznych konieczne jest zastosowanie metod numerycznych przybliżających rozwiązanie problemu.

Modele matematyczne można również podzielić na modele dedukcyjne oraz indukcyjne. Modele dedukcyjne tworzy się przy pomocy reguł wnioskowania dedukcyjnego na podstawie pewnej koncepcji przyjętej a priori. W tym przypadku dane empiryczne stosuje się do weryfikacji zbudowanego modelu. Natomiast modele indukcyjne tworzy się bezpośrednio na podstawie danych empirycznych, które umożliwiają dokonanie indukcyjnych uogólnień i stawianie hipotez. Podstawą struktury takich modeli jest najczęściej (określona przez badacza) pewna zależność matematyczna. Jednak przyjęcie takiego modelu musi zostać poprzedzone oceną wielkości błędu, z jakim opisuje on strukturę badanego systemu lub zjawiska.

Z kolei modele schematyczne stosowane są najczęściej jako dwuwymiarowe modele rysunkowe wykonane w określonej skali. Stanowią one graficzne lub symboliczne sposoby przedstawienia systemów, procesów lub obiektów, które identyfikują występujące w nich elementy, relacje i zależności.

Można wyróżnić również podział modeli ze względu na szereg różnych kryteriów, tj. zależność od czasu, rodzaj modelowanego systemu, zmienność cech itp. Ze względu na zależność od czasu wyróżnia się modele:

- statyczne – zmienne opisujące model nie zależą od czasu,
- dynamiczne – zmienne opisujące model zależą od czasu.

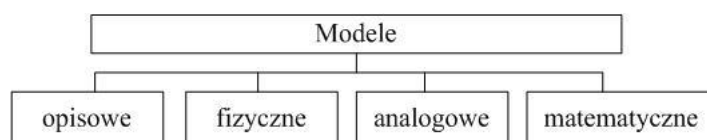
Ze względu na rodzaj modelowanego systemu wyróżnia się modele:

- dyskretne – zmienne stanu zmieniają się jedynie w określonych momentach,
- ciągłe – zmienne stanu zmieniają się w sposób ciągły w czasie.

Z kolei uwzględniając sposób reprezentacji zmiennych, modele można również podzielić na:

- stochastyczne – relacje między obiektami nie są jednoznacznie określone i przyjmują charakter losowy,
- deterministyczne – relacje między obiektami oraz atrybuty obiektów są ściśle określone (zmienne są zdefiniowane lub określone według zadanych funkcji matematycznych).

Modele można również podzielić ze względu na charakter uproszczenia rzeczywistego obiektu (rys. 29) [48]. W tym kontekście można wyróżnić modele: opisowe (opisujące właściwości obiektu w sposób werbalny), fizyczne (przedstawiające działanie obiektu rzeczywistego z wykorzystaniem wielkości fizycznych), analogowe (opisujące schematy połączeń różnych układów oparte na danych technicznych rzeczywistych urządzeń), matematyczne (przedstawiający relacje ilościowe między wielkościami związanymi z obiektem) [48].



Rys. 29. Podział modeli ze względu na stopień zastosowanego uproszczenia [48]

W literaturze można wyróżnić również szereg innych klasyfikacji modeli ze względu na bardziej szczegółowe kryteria podziału, tj. docelowe zastosowanie, język opisu oraz obszar zastosowania, które zostały przedstawione w tabeli 10.

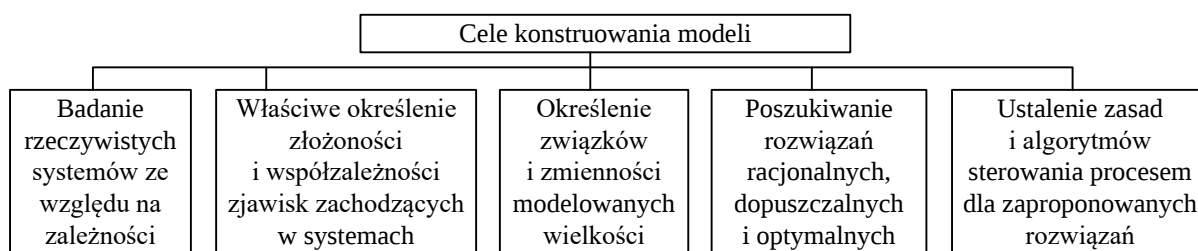
Tabela 10. Klasyfikacja modeli ze względu na wybrane kryteria podziału [48]

Kryterium podziału	Rodzaje modeli
Ze względu na cel przeznaczenia	Modele wyjaśniające
	Modele oceny
	Modele decyzyjne
	Modele prognostyczne
Ze względu na język opisu	Modele werbalne
	Modele logiczne
	Modele matematyczne
Ze względu na obszar zastosowania	Modele obiektów punktowych
	Modele obiektów liniowych

Najważniejsze cele budowania modeli zostały przedstawione w uproszczeniu na rysunku 30 i zalicza się do nich m.in.:

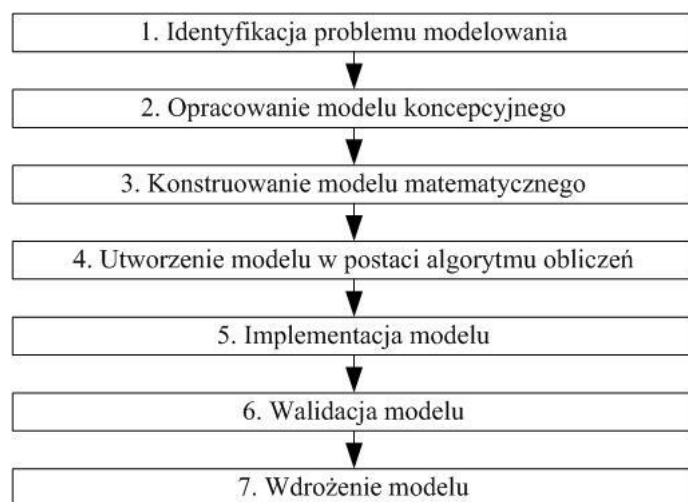
- badanie rzeczywistych systemów biorąc pod uwagę zależności jakie zachodzą pomiędzy ich elementami oraz właściwościami,
- ustalenie zależności i współzależności zjawisk zachodzących w systemie oraz z jego otoczeniem,

- określenie związków pomiędzy modelowanymi wielkościami oraz ich zmienności, w celu uzyskania informacji o obiekcie rzeczywistym,
- poszukiwanie rozwiązań racjonalnych (przez dobranie najkorzystniejszych parametrów), dopuszczalnych (spełniających określone minimalne ograniczenia) oraz optymalnych (spełniających określone ograniczenia modelu oraz zapewniające maksymalną ocenę jakości zaproponowanego rozwiązania),
- określenie zasad i algorytmów sterowania procesem, w celu spełnienia określonych ograniczeń dla zaproponowanych rozwiązań.



Rys. 30. Cele konstruowania modeli [48]

Tworzenie modeli składa się z szeregu etapów przedstawionych graficznie na rysunku 31 oraz polegających na realizacji pewnego określonego ciągu czynności. Budowanie modelu jest każdorazowo ukierunkowane na rozwiązywanie konkretnego problemu.



Rys. 31. Etapy budowy modeli [48]

Przykładowa procedura konstruowania modeli obejmuje następujące etapy [48]:

- identyfikacja komponentów systemu i występujących powiązań pomiędzy nimi – identyfikacja odpowiednich elementów modelowanego systemu wraz z ustaleniem zależności występujących pomiędzy nimi. W tym kroku należy określić listę elementów sterujących, połączeń i aspektów, które są zidentyfikowane w porównaniu z tymi, które wymagają dalszego zbadania,

- opracowanie modelu koncepcyjnego – etap polega na transformacji zbioru założeń do postaci modelu koncepcyjnego, określeniu relacji pomiędzy zidentyfikowanymi elementami oraz ustalenie niezbędnych cech do opisu danego systemu lub obiektu,
- budowa modelu matematycznego – tzw. finalizacja modelu, która pozwala na tworzenie algorytmów obliczeń,
- algorytmizacja modelu – polega na przetransferowaniu modelu formalnego w strukturę tzw. generatora stanu obiektu. W etapie tym opracowane zostają formuły obliczeniowe modelowanego obiektu lub systemu,
- zaimplementowanie modelu – etap składający się z działań zmierzających do zamiany modelu algorytmicznego na model informatyczny (aplikacji komputerowej z zaimplementowanymi algorytmami obliczeniowymi),
- walidacja modelu – polega na ocenie dopasowania się modelu do rozwiązania danego problemu. Model należy zweryfikować przez porównanie jego wyników i zachowań z rzeczywistym systemem (etap walidacji modelu). Jeśli wyniki uzyskane przy pomocy modelu odpowiadają wynikom rzeczywistym, należy powrócić do poprzednich kroków w celu udoskonalenia modelu. Omawiany proces jest wykonywany iteracyjnie w pętli aż do uzyskania akceptowalnego modelu (etap nauki z procesu),
- wdrożenie modelu – polega na przeprowadzeniu eksperymentów w warunkach rzeczywistych, w celu określenia zakresu, warunków stosowalności i możliwości posługiwania się modelem.

Modele mogą przyjmować różne formy, od modeli matematycznych, aż po zaawansowane symulacje komputerowe. Współczesne techniki modelowania są stosowane m.in. w kontekście analizy i prognozowania emisji gazów cieplarnianych, umożliwiając tworzenie dokładnych modeli emisji CO₂.

Podsumowując, modelowanie to iteracyjny proces, na który składa się szereg wymienionych i opisanych kroków, którego efektem końcowym jest uproszczona reprezentacja zapewniająca użyteczny opis badanego, rzeczywistego zjawiska. Uproszczona struktura modelu w stosunku do obiektu rzeczywistego powoduje, że jego badanie dostarcza informacji ważnych i przydatnych do celów badawczych.

4.2. Algorytm badania oraz wstępna analiza danych

4.2.1. Charakterystyka zmiennych wykorzystanych w badaniu

Do budowy modelu zastosowano zmienne niezależne zarejestrowane podczas badań:

- prędkość obrotową silnika,
- obciążenie silnika,
- moc silnika,
- prędkość pojazdu,
- moment obrotowy silnika,
- położenie przepustnicy.

Zmienną zależną stanowi wartość chwilowej emisji CO₂ w samochodzie. W dalszej części podrozdziału wymienione wyżej zmienne zostaną szczegółowo scharakteryzowane.

Prędkość obrotowa silnika

Prędkość obrotowa stanowi jeden z podstawowych wskaźników pracy silnika, który określa wartość chwilowej liczby obrotów wału korbowego silnika w określonej jednostce czasu (1):

$$n = \frac{\Delta r}{\Delta t}, \quad (1)$$

gdzie:

n – prędkość obrotowa silnika,

Δr – zmiana liczby obrotów wału silnika,

Δt – czas.

Z praktycznych powodów, omawianą wielkość podaje się często w obr/min lub obr/s. W praktyce inżynierskiej często spotyka się pierwsze z wymienionych podejść. Z uwagi na to, iż obecne silniki samochodowe należą do silników szybkoobrotowych (dla ZI $n > 2800$ obr/min, dla ZS $n > 1200$ obr/min), omawiany parametr jest kluczowy w ocenie stanu pracy silnika.

Prędkość obrotowa silnika wzrasta wraz z naciśnięciem pedału przyspieszenia, co z kolei zwiększa moc silnika. Prędkość obrotową silnika mierzy się przy pomocy tachometrów, które zapewniają bezpośredni odczyt prędkości obrotowej wału korbowego w obr/min. Parametr można mierzyć niezależnie od ruchu pojazdu, ponieważ stanowi on właściwość samego silnika. Zalecany zakres prędkości obrotowej silnika różni się w zależności od konstrukcji silnika, warunków jazdy oraz innych parametrów, np. obciążenia. Silniki są projektowane w ten sposób, aby działały najskuteczniej w określonym przez producenta

zakresie obrotów. Kontrolowanie prędkości obrotowej silnika ma kluczowe znaczenie dla zużycia paliwa, a w konsekwencji na emisję CO₂.

Obciążenie silnika

Obciążenie silnika określa wartość pracy, jaką w danej chwili wykonuje silnik, co jest bezpośrednio związane z wartościami zużywanego przez silnik powietrza i paliwa. Przy danej prędkości obrotowej silnika w danej chwili, można scharakteryzować obciążenie silnika jego mocą, które może być wyrażane w jednostkach absolutnych [kW] lub względnych [%]. Względne obciążenie definiuje się jako stosunek chwilowej mocy silnika do jego maksymalnej mocy (2):

$$L_{\%} = \frac{P}{P_{max}} \cdot 100\%, \quad (2)$$

gdzie:

$L_{\%}$ - względne obciążenie silnika,

P – moc silnika,

P_{max} – maksymalna moc silnika.

Dokładna metoda obliczania tego parametru może się różnić w zależności od różnych systemów zarządzania/sterowania silnikiem. Najczęściej wykorzystywane są dane z czujników, tj. ciśnienie bezwzględne w kolektorze dolotowym, masowy przepływ powietrza, położenie przepustnicy, prędkość obrotowa silnika czy temperatura powietrza otoczenia. Następnie ECU oblicza omawiany parametr w oparciu o zaimplementowany model matematyczny, określony przez producenta pojazdu. Obliczone względne obciążenie silnika osiąga 100% przy całkowicie otwartej przepustnicy, niezależnie od wysokości terenu, temperatury otoczenia i prędkości obrotowej silnika. Parametr wskazuje procent dostępnego momentu obrotowego wytwarzanego przez silnik. Monitorowanie względnego obciążenia silnika jest źródłem informacji na temat warunków pracy i wydajności silnika, natomiast sam parametr jest kluczowy w zarządzaniu jego pracą.

Moc silnika

Moc silnika jest wskaźnikiem pracy silnika opisującym zdolność silnika do wykonywania pracy. Moc w układach fizycznych stanowi skalarną wielkość fizyczną, którą definiuje się jako iloraz wykonanej pracy w danej jednostce czasu (3):

$$P = \frac{W}{t}, \quad (3)$$

gdzie:

P – moc,

W – praca,

t – czas.

Korzystając z przedstawionej definicji, chwilową moc w przypadku silnika spalinowego można przedstawić w następujący sposób (4):

$$P = M \cdot n, \quad (4)$$

gdzie:

P – moc silnika,

M – moment obrotowy,

n – prędkość obrotowa silnika.

Moc silnika podaje się najczęściej w koniach mechanicznych KM (historycznie) lub zgodnie z układem SI w kilowatach kW. Pomiaru mocy silnika dokonuje się najczęściej w wyspecjalizowanych laboratoriach badawczych lub w stacjach diagnostycznych z wykorzystaniem hamowni podwoziowej, która umożliwia symulację warunków drogowych. Występuje wiele odmian tego parametru w odniesieniu do pojazdów samochodowych, tj. moc na kołach: użyteczna, nominalna, krótkotrwała, trwała itd.

Prędkość pojazdu

W modelowaniu ruchu pojazdów samochodowych często przyjmuje się założenie, że ruch pojazdu opisywany jest przez ruch wybranego punktu materialnego, zwykle środka geometrycznego pojazdu. W ten sposób ruch pojazdu można sklasyfikować jako krzywoliniowy z ciągle zmieniającą się wartością promienia krzywizny [94]. W omawianym kontekście ruch prostoliniowy występuje, gdy wartość promienia krzywizny dąży do nieskończoności. Prędkość pojazdu stanowi w tym kontekście kinematyczny parametr ruchu pojazdu, przedstawiany w postaci prędkości liniowej (stycznej do toru jazdy), którą można określić w następujący sposób (5):

$$v = \frac{\Delta s}{\Delta t}, \quad (5)$$

gdzie:

v – prędkość pojazdu,

Δs – przebyta droga,

Δt – czas.

W kontekście ruchu samochodu prędkość pojazdu to odległość pokonywana w jednostce czasu. Parametr ten charakteryzuje się dużą zmiennością, w związku z czym do jego opisu często podaje się jego wartość średnią.

Prędkość pojazdu jest jednym z parametrów dostępnych do odczytu w interfejsie OBD. Jest określana pośrednio przez pomiar prędkości obrotowej kół realizowany przez tachometry. Ten sposób wymaga jednak znajomości dodatkowych parametrów związanych z ogumieniem, tj. średnica opony, ciśnienie itd. Ponadto, zgodnie z obowiązującym prawem, prędkościomierz w samochodzie może wskazywać prędkość wyższą od rzeczywistej – maksymalny dozwolony błąd pomiaru stanowi sumę 3% prędkości rzeczywistej oraz nie mniej niż 3 km/h [97].

Prędkość pojazdu można również określić w inny sposób, na przykład wykorzystując odbiornik GNSS telefonu komórkowego. Badania wskazują, iż to rozwiązanie posiada szereg zalet w stosunku do opisanej wcześniej konwencjonalnej metody pomiaru prędkości pojazdu. Pomiar prędkości przy pomocy GNSS jest całkowicie niezależny od pojazdu oraz warunków jazdy, ponieważ nie jest zakłócany przez takie czynniki jak np. poślizg kół, nachylenie terenu czy wartości wskazań czujników pokładowych w samochodzie. Kolejnym czynnikiem jest brak regulacji prawnych, sztucznie zawyżających wskazania prędkości. Wartości prędkości zależą od dostępnej w danej chwili liczby oraz konstelacji sygnałów z satelitów, co sprawia, iż wskazania te są bliższe wartościom rzeczywistym, niż wskazania prędkościomierza w samochodzie.

Prędkość pojazdu stanowi parametr krytyczny dla osiągow silnika oraz zużycia paliwa. Parametr wpływa bezpośrednio na warunki pracy silnika i moc przekazywaną na koła. Uzyskanie wyższych prędkości pojazdu zazwyczaj wymaga większej mocy silnika (moc silnika zależy od trzeciej potęgi prędkości pojazdu), co wpływa na zużycie paliwa i emisję CO₂. Ponadto, prędkość pojazdu jest kluczowym czynnikiem kształtującym bezpieczną jazdę, ponieważ wpływa na zdolność kierowcy do panowania nad pojazdem i odległość niezbędną do jego zatrzymania.

Moment obrotowy silnika

Moment obrotowy w samochodzie spalinowym odnosi się do momentu siły generowanej przez wał korbowy silnika. Parametr ten (6), definiuje się jako iloczyn siły stycznej F_T powodującej obrót wału korbowego oraz promień wykorbienia wału korbowego r (rys. 32):

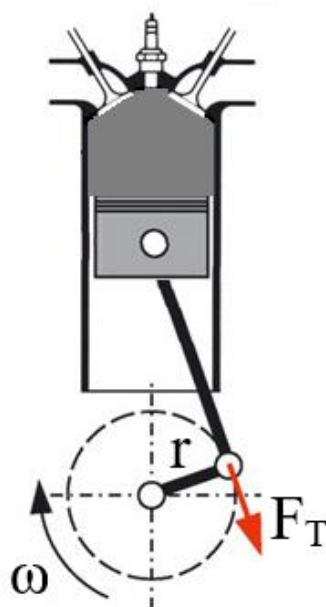
$$M = F_T \cdot r, \quad (6)$$

gdzie:

M – moment obrotowy,

F_T – siła styczna,

r – promień wykorbienia wału korbowego.



Rys. 32. Schemat powstawania momentu obrotowego w silniku spalinowym [25]

Jednostką momentu obrotowego jest Nm. Moment obrotowy jest wielkością fizyczną wektorową, która zmienia swoją wartość podczas jednego obrotu wału korbowego, natomiast jego wartość chwilowa zależy od wielu czynników, tj. ciśnienie gazów w cylindrze oraz kąt obrotu wału korbowego. Z uwagi na konstrukcję elementów silnika spalinowego, moment obrotowy na kole zamachowym stanowi sumę momentów sił ze wszystkich występujących cylindrów silnika [38].

Natomiast z wymienionych powodów praktycznych mierzona wartość momentu obrotowego silnika jest wartością średnią momentu siły wytwarzanej przy określonej prędkości obrotowej dla jednego cyklu pracy silnika. Najczęściej moment obrotowy silnika jest określany na hamulcu stanowiska dynamometrycznego. Parametr ten stanowi kluczowy element w procesie wytwarzania energii przez silnik i odgrywa znaczącą rolę w kształtowaniu osiązków pojazdu.

Położenie przepustnicy (dla silników ZI)

Położenie przepustnicy w samochodzie odpowiada kątowi otwarcia przepustnicy, która reguluje ilość powietrza doprowadzanego do komór spalania silnika. Pomiar położenia przepustnicy stanowi najczęściej pomiar względny, podawany w procentach, określany zgodnie z zależnością (7):

$$\varphi_{\%} = \frac{\varphi}{\varphi_{max}} \cdot 100\%, \quad (7)$$

gdzie:

$\varphi_{\%}$ – względne położenie przepustnicy,

φ – kąt otwarcia przepustnicy,

φ_{max} – maksymalny kąt otwarcia przepustnicy.

Położenie przepustnicy jest mierzone za pomocą czujników wykrywających kąt nachylenia przepustnicy, natomiast w samochodach z elektronicznym sterowaniem przepustnicą, ECU steruje pracą omawianego elementu. Położenie przepustnicy ma kluczowe znaczenie w kontekście sterowania pracą silnika, ponieważ bezpośrednio wpływa na jego moc wyjściową. Przez regulację położenia przepustnicy możliwa jest nie tylko regulacja osiągow silnika, ale również zużycia paliwa czy emisji spalin.

Emisja CO₂

Emisja CO₂ w samochodach spalinowych dotyczy ilości dwutlenku węgla wytwarzanego jako produkt uboczny spalania paliwa w silniku. Jednostką emisji CO₂ jest g/km dla pojazdów typu LDV. W niniejszym badaniu emisja CO₂ stanowi zmienną wyjściową, a jej chwilowa wartość jest możliwa do odczytu z ECU silnika dzięki zastosowaniu interfejsu OBD. Układ kontroli emisji spalin w pojazdach nie jest jednak wyposażony w czujnik CO₂, natomiast na podstawie zużycia paliwa przez pojazd można obliczyć równoważną emisję CO₂ (zużycie paliwa o wartości 1 dm³/100 km powoduje emisję CO₂: dla benzyny – 23,3 g/km a dla oleju napędowego – 26,3 g/km). Istnieje wiele sposobów szacowania emisji CO₂, co zostało opisane w pracy [80], jednak najczęściej wykorzystuje się w tym celu natężenie przepływu paliwa w silniku.

Zgodnie z ideą działania układu wtrysku paliwa, ECU silnika w czasie rzeczywistym monitoruje i rejestruje wartość paliwa trafiającego do każdego cylindra i na tej podstawie oblicza natężenie przepływu paliwa, którego wartość chwilowa jest przesyłana do magistrali komunikacyjnej pojazdu, co umożliwia następnie dostęp do niej jako jednego z parametrów diagnostycznych w systemie OBD. Omawiany parametr można zapisać zgodnie z równaniem (8) [42]:

$$\dot{m}_p = \frac{q \cdot n \cdot z}{30 \cdot \tau}, \quad (8)$$

gdzie:

$\dot{m}_p$ – natężenie przepływu paliwa,

q – masa wtryskiwanego paliwa do cylindra na jeden cykl,

n – prędkością obrotową silnika,

z – liczba cylindrów silnika,

τ – liczba obrotów wału korbowego silnika na cykl.

Na podstawie wartości chwilowego natężenia przepływu paliwa, można określić natężenie emisji CO₂ (9) [42]:

$$\dot{m}_{CO_2} = k_{CO_2} \cdot \dot{m}_p, \quad (9)$$

gdzie:

$\dot{m}_{CO_2}$ – natężenie emisji CO₂,

k_{CO_2} – współczynnik emisji CO₂ dla paliwa kopalnego.

Współczynnik emisji CO₂ jest parametrem reprezentującym ilość emitowanego CO₂ podczas zużycia danego rodzaju paliwa. Współczynnik ten jest najczęściej podawany przez producenta pojazdu. Wartość jednostkowej emisji CO₂ można określić przy pomocy zależności (10) [42]:

$$m_i = \int_{t_0}^{t_k} \dot{m}_{CO_2} dt, \quad (10)$$

gdzie:

m_i – wartość jednostkowej emisji CO₂,

t_0, t_k – czas rozpoczęcia i zakończenia pojedynczego cyklu pomiarowego.

Ostatecznie wartość chwilowego wskaźnika emisji CO₂ w samochodzie można określić przy pomocy zależności (11) [42]:

$$m_{CO_2} = \frac{m_i}{s_i}, \quad (11)$$

gdzie:

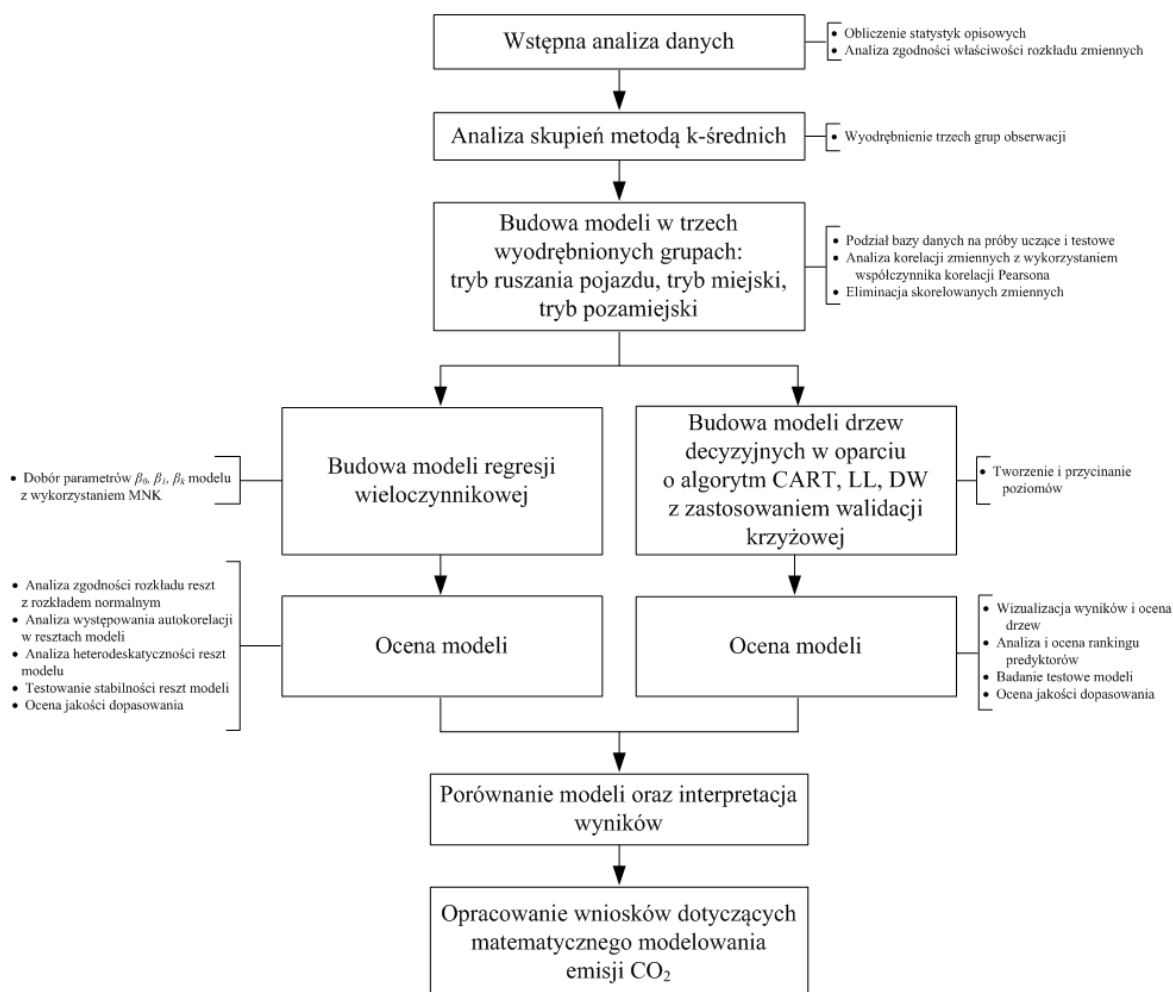
m_{CO_2} – wskaźnik emisji CO₂,

s_i – przebyta droga w jednostkowym cyklu pomiarowym.

Obliczanie emisji CO₂ w systemach OBD stanowi zatem skomplikowany proces, na który wpływa wiele parametrów, m.in. natężenie przepływu paliwa, rodzaj paliwa, dane techniczne silnika i współczynnik emisji CO₂.

Podsumowując, wiedza dotycząca wpływu wymienionych parametrów pozyskanych z interfejsów diagnostycznych OBD, może być skutecznym wsparciem w kreowaniu stylu jazdy kierowcy pod kątem minimalizacji emisji CO₂. Zależności te mogą być stwierdzone za pomocą modeli matematycznych opisujących emisję CO₂ w pojazdach samochodowych. Ponadto interfejsy diagnostyczne OBD mogą stanowić ważne narzędzie w dążeniu do zapewnienia zrównoważonego rozwoju w sektorze transportu i wsparcia systemów zarządzania środowiskowego, umożliwiając podejmowanie świadomych decyzji zarówno na poziomie indywidualnego użytkownika, jak i korporacyjnym czy rządowym. Rozsądne podejście kierowcy do prowadzenia samochodu i utrzymania go w dobrym stanie technicznym mogą znacząco przyczynić się do redukcji śladu węglowego związanego z transportem drogowym.

Realizacja badań w niniejszej dysertacji związana z modelowaniem emisji CO₂ do środowiska przebiegała zgodnie z algorytmem przedstawionym na rysunku 33.



Rys. 33. Algorytm modelowania emisji CO₂ w oparciu o dane rzeczywiste

Badanie podzielono na trzy główne etapy, związane ze wstępną analizą danych, analizą skupień, budową i oceną modeli matematycznych. Ze względu na charakter zmiennej objaśnianej (przyjmującej wartości liczbowe) zdecydowano się na modele regresyjne, w tym regresję wieloraką oraz drzewa decyzyjne z zastosowaniem algorytmu CART, lasy losowe oraz drzewa wzmacniane. Analizę danych wykonano z wykorzystaniem oprogramowania *Statistica 13.1*. W pierwszym etapie przeprowadzono wstępną analizę danych z wykorzystaniem podstawowych statystyk opisowych oraz ocenę przebiegu zmiennych w odniesieniu do emisji CO₂.

Badano:

- średnią arytmetyczną (12):

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (12)$$

gdzie:

x_i – obserwacje w zbiorze danych,

n – liczba obserwacji.

- medianę (13):

$$Me = \begin{cases} x_{\frac{n+1}{2}}, & \text{jeśli } n \text{ jest nieparzyste} \\ \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}, & \text{jeśli } n \text{ jest parzyste} \end{cases}, \quad (13)$$

- odchylenie standardowe (14):

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}, \quad (14)$$

- kurtozę (15):

$$K = \frac{\mu_4}{\sigma^4}, \quad (15)$$

gdzie:

μ_4 – wartość czwartego momentu centralnego.

- skośność (16):

$$A = \frac{\mu_3}{\sigma^3}, \quad (16)$$

gdzie:

μ_3 – wartość trzeciego momentu centralnego.

Obliczenie powyższych charakterystyk umożliwiło ocenę występowania obserwacji odstających oraz wstępną ocenę przebiegu rozkładu zmiennych. Przyjmuje się, że jeżeli kurtoza znajduje się w przedziale (-2 do 2) to rozkład jest zbliżony do normalnego, natomiast skośność z przedziału (-1,5 do 1,5) pozwala wnioskować o symetryczności rozkładu.

Następnie dokonano sprawdzenia własności rozkładu badanych zmiennych. Do weryfikacji powyższego wykorzystano test Kołmogorowa–Smirnowa. W teście tym hipoteza zerowa zakłada:

H_0 : nie ma różnicy między rozkładem normalnym a rozkładem z próby,

natomiast alternatywna:

H_1 : rozkład z próby istotnie różni się od rozkładu normalnego.

Test Kołmogorowa–Smirnowa wykorzystuje odległość pomiędzy dystrybuantą empiryczną a teoretyczną rozkładu normalnego $N(\mu, \sigma^2)$. Powinien być stosowany względem licznych prób. W przypadku, gdy parametry rozkładu nie są znane, to wyznaczamy oceny estymatorów wartości średniej μ i odchylenia standardowego σ na podstawie próby. Dla

uporządkowanej realizacji próby $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ definiujemy dystrybuantę empiryczną w postaci (17) [13]:

$$F_n(x) = \begin{cases} 0, & \text{dla } x < x_{(1)}, \\ \frac{i}{n}, & \text{dla } x_{(i)} \leq x < x_{(i+1)}, 1 \leq i < n, \\ 1, & \text{dla } x \geq x_{(n)}. \end{cases} \quad (17)$$

Dystrybuanta teoretyczna rozkładu normalnego wyraża się wzorem (18):

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds, \quad (18)$$

wartość statystyki testowej D_n wyznaczana jest jako (19):

$$D_n = \max_{1 \leq i \leq n} |F_n(x_{(i)}) - F(x_{(i)})|. \quad (19)$$

Statystyka testu D_n ma rozkład Kołmogorowa-Smirnowa. Dla poziomu istotności α oraz liczebności próby n wyznaczamy wartość krytyczną $d_n(1 - \alpha)$. Jeżeli $D_n < d_n(1 - \alpha)$ to na poziomie istotności α nie podstaw do odrzucenia H_0 . W przypadku gdy $D_n \geq d_n(1 - \alpha)$ to H_0 odrzucamy na korzyść H_1 .

Hipotezy możemy również zweryfikować za pomocą prawdopodobieństwa testowego - wartości $p = P(K \geq D_n)$, gdzie K oznacza zmienną losową o rozkładzie Kolmogorowa. Jeżeli wartość $p < \alpha$ to hipotezę H_0 odrzucamy na korzyść hipotezy alternatywnej H_0 .

W kolejnym etapie dokonano wizualizacji przebiegu badanych zmiennych w odniesieniu do emisji CO₂ wykorzystując w tym celu wykresy liniowe. Pozwoliło to stwierdzić, że w analizowanym zbiorze zauważalne są różniące się istotnie grupy, które w ramach dalszych obliczeń zostały wyodrębnione z zastosowaniem metody k-średnich.

Metoda k -średnich jest jedną z metod stosowanych do analizy skupień, umożliwiającą grupowanie n obiektów, które opisane są za pomocą p cech, w K niepustych, rozłącznych i możliwie jednorodnych grupach – skupieniach. Celem takiej analizy jest wykrycie „naturalnych” skupień, które dają się interpretować. W ramach metody zakłada się istnienie n obserwacji, które można przedstawić wektorami $x_1, x_2, \dots, x_n$ o wymiarze m . Omawiany algorytm dokonuje podziału tych obserwacji na pewną liczbę K zbiorów $S_1, S_2, \dots, S_K$, gdzie $K < n$ tak, aby zminimalizować przyjętą funkcję celu J (20):

$$J = \sum_{k=1}^K \sum_{x_i \in S_K} \|x_i - c_k\|^2, \quad (20)$$

gdzie:

c_k – średnia wartość zmiennych dla obiektów należących do skupienia S_K ,

x_i – wartość zmiennej x dla obiektu należącego do skupienia S_K ,

K – liczba skupień.

Pomocna w ustaleniu optymalnej liczby skupień jest między innymi wykorzystana w niniejszej pracy metoda łokcia. Metoda ta bazuje na graficznej reprezentacji znajdowania optymalnej liczby K (skupień) w klasteryzacji k -średnich. W ramach tej metody poszukiwana jest wartość sumy błędów kwadratowych (SSE), tj. sumy kwadratowej odległości między punktami w klastrze a centroidem klastra. Na osi Y obrazowana jest wartość SSE , która odpowiada różnym wartościom K znajdującym się na osi X . W ramach metody dąży się do minimalizacji łącznej sumy wewnątrzklustrowej sumy kwadratów. Dodatkowo, może ona być wsparta wizualną inspekcją wykresu, polegającą na poszukiwaniu takiej wartości K , dla której tworzony jest tzw. „punkt łokciowy”.

W rozwiązaniu problemu k -średnich stosuje się dwie metody: bezpośrednią oraz skumulowaną. W pierwszej z nich, zwanej również metodą on-line, aktualizacja centrów dokonywana jest iteracyjnie wraz z każdorazowym przydzieleniem wektora x pochodzącego ze zbioru danych uczących. Wersja skumulowana (off-line) zakłada z kolei, iż centra będą aktualizowane jednocześnie po podaniu pełnego wektora uczącego. Metoda bezpośrednia jest stosowana w sytuacji, gdy układ otrzymuje dane wejściowe sekwencyjnie lub w czasie rzeczywistym, natomiast w przypadku, gdy cały wektor jest już znany stosuje się wersję skumulowaną.

W metodzie skumulowanej dane są wszystkie wektory uczące x i każdy z nich przyporządkowany jest do odpowiedniego centrum według metryki odległościowej. Zbiór wektorów przypisanych do określonego centrum tworzy grupę, którego nowe centrum można określić zgodnie z poniższą zależnością (21):

$$c_i(k+1) = \frac{1}{n_i} \sum_{i=1}^{n_i} x_i, \quad (21)$$

gdzie: n_i – liczba wektorów x przyporządkowanych w k -tym cyklu do i -tego centrum.

W przypadku tej metody aktualizacja wartości wszystkich centrów odbywa się równolegle, a proces określenia ich pozycji jest powtarzany wielokrotnie, aż do ustalenia się końcowych wartości. Ponadto ważnym zagadaniem jest również wstępny wybór centrów, najczęściej odbywający się losowo przez wykorzystanie rozkładu równomiernego składowych wektora [82].

W kolejnym kroku dokonano sprawdzenia zgodności rozkładów emisji CO_2 w wyodrębnionych grupach. W tym celu zastosowanie znajduje analiza wariancji (ANOVA). Jednoczynnikowa analiza wariancji znajduje zastosowanie wówczas, gdy rozważamy

zależność pomiędzy zmienną ilościową Y , a jakościową zmienną objaśniającą X , zwaną w literaturze czynnikiem klasyfikującym lub też zmienną grupującą. Zmienna objaśniająca X przyjmuje k możliwych realizacji (tzn. liczba grup obserwacji wynosi k , lub analizujemy k -populacji) [13].

Dla każdej populacji losujemy próbę o realizacji $\{y_{ij}\}_{1 \leq j \leq n_i}$, gdzie y_{ij} oznacza j -tą obserwację w i -tej grupie (dla i -tej populacji), dla i -tej grupy liczba obserwacji wynosi n_i , $1 \leq i \leq k$.

Model analizy wariancji można przedstawić w poniższy sposób (22):

$$y_{ij} \sim N(\mu_i, \sigma^2), \quad (22)$$

dla $1 \leq i \leq k$, $1 \leq j \leq n_i$, całkowita liczba obserwacji $n = \sum_{i=1}^k n_i$, natomiast μ_i – oznacza średnią dla i -tej grupy.

Główne zastosowanie analizy wariancji polega na weryfikacji hipotezy o równości średnich w grupach na podstawie wyników próby. Na poziomie istotności α tworzymy hipotezę roboczą:

$H_0: \mu_1 = \mu_2 = \dots = \mu_k$ – średnie w grupach są równe,

wobec hipotezy alternatywnej:

H_1 : istnieją grupy $i, j (1 \leq i, j \leq k)$ takie, że $\mu_i \neq \mu_j$ – zatem istnieją przynajmniej dwie grupy, dla których średnie istotnie się różnią.

W analizie wariancji przyjmuje się model opisujący zmienną zależną ze względu na zmienną grupującą. Dla i -tej grupy (23):

$$y_{ij} = \mu + a_i + \varepsilon_{ij}, \quad (23)$$

gdzie:

a_i – różnica pomiędzy średnią μ_i w grupie i -tej grupy a średnią globalną μ (średnia dla wszystkich populacji),

$\{\varepsilon_{ij}\}_{1 \leq j \leq n_i}$ – elementy ciągu są zmiennymi losowymi o rozkładzie normalnym $\varepsilon_{ij} \sim N(0, \sigma^2)$ oraz $cov(\varepsilon_{il}, \varepsilon_{is}) = 0$ dla $l \neq s$, $1 \leq l, s \leq n_i$.

a_i – efekt (parametr) informuje o różnicy wartości oczekiwanej zmiennej objaśnianej w i -tej grupie od średniej dla wszystkich populacji: $a_i = E y_{ij} - \mu$ dla $1 \leq j \leq n_i$.

Zatem w analizie wariancji weryfikowana jest hipoteza (24):

$$H_0: a_1 = a_2 = \dots = a_k = 0, \quad (24)$$

która jest równoważna hipotezie roboczej (25):

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k. \quad (25)$$

Wielkość (26):

$$\bar{y} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}, \quad (26)$$

oznacza estymator średniej dla wszystkich populacji (grup), natomiast:

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}, \quad (27)$$

oznacza średnią dla i -tej grupy, $i = 1, 2, \dots, k$.

W analizie wariancji wyznaczamy zmienność całkowitą (suma kwadratów różnic od średniej grupowej $\bar{y}$, SST - sum of squares total).

$$\begin{aligned} SST &= \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i + \bar{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2. \end{aligned} \quad (28)$$

Wielkość:

$$SSW = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2, \quad (29)$$

nazywamy zmiennością wewnątrzgrupową (z ang. sum of squares within groups), natomiast:

$$SSB = \sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2, \quad (30)$$

nazywamy zmiennością międzygrupową (z ang. sum of squares between groups).

Im bardziej zmienność wewnątrzgrupowa SSW różni się od zmienności całkowitej SST , tym większa jest wartość zmienności międzygrupowej SSB . Hipotezę zerową odrzucamy, gdy SSB jest dostatecznie duże w stosunku do SSW . Statystyka testowa:

$$F = \frac{SSB}{SSW} \frac{n - k}{k - 1}, \quad (31)$$

ma rozkład Fishera – Snedecora o $(k - 1)$ i $(n - k)$ stopniach swobody.

Obszar krytyczny jest postaci:

$$W = [F_{1-\alpha}(k - 1, n - k), +\infty), \quad (32)$$

Jeżeli $F \notin W$ to na poziomie istotności nie ma podstaw do odrzucania hipotezy H_0 , możemy przyjąć, że średnie w grupach są równe bądź nieistotnie się różnią. Jeżeli $F \in W$ to na poziomie istotności odrzucamy hipotezę H_0 na korzyść H_1 , średnie w grupach różnią się istotnie, to znaczy, że istnieją co najmniej takie dwie grupy, dla których różnica średnich istotnie różni się od zera [13].

Stosowanie analizy wariancji ANOVA wymaga spełnienia założeń:

- zmienna objaśniana ma rozkład normalny w populacjach (grupach),
- potwierdzona jest jednorodność wariancji w populacjach (grupach).

W rozprawie, pierwsze założenie weryfikowane jest przez test Kołomogorova–Smirnova. Natomiast jednorodność wariancji sprawdzana jest z zastosowaniem testu Levene’a.

W ramach powyższego testu sprawdzane są następujące hipotezy:

H_0 : Wariancja zmiennej zależnej w grupach czynnika jest jednakowa: $s^2_1 = s^2_2 = \dots = s^2_k$.

H_1 : Wariancja zmiennej zależnej w grupach czynnika różni się.

W teście Levene’a dla każdej realizacji x_{ij} (i -ta grupa, j -ta realizacja dla $1 \leq j \leq n_i$) wyznaczamy wartości bezwzględne różnic względem średniej lub mediany:

$$y_{ij} = \begin{cases} |x_{ij} - \bar{x}_i|, & \bar{x}_i - \text{średnia w } i - \text{tej grupie,} \\ |x_{ij} - \hat{x}_i|, & \hat{x}_i - \text{mediana w } i - \text{tej grupie,} \end{cases} \quad (33)$$

oraz wyznaczamy:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}, \quad (34)$$

gdzie:

$$n = \sum_{i=1}^k n_i, \quad (35)$$

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}. \quad (36)$$

Statystyka testowa (37):

$$F = \frac{n - k}{k - 1} \frac{\sum_{j=1}^k n_i (\bar{y}_i - \bar{y})^2}{\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}, \quad (37)$$

ma rozkład Fishera-Snedecora o $(k - 1, n - k)$ stopniach swobody. Obszar krytyczny jest postaci (38):

$$W = [F_{1-\alpha}(k - 1, n - k), +\infty), \quad (38)$$

Jeżeli $F \in W$ to na poziomie istotności α hipotezę zerową H_0 odrzucamy na korzyść H_1 (wariancje w grupach różnią się istotnie). Jeżeli $F \notin W$ to na poziomie istotności α nie ma podstaw do odrzucenia hipotezy H_0 .

W przypadku braku potwierdzenia założeń analizy wariancji, do oceny istotności różnic pomiędzy średnimi stosuje się test Kruskala-Wallisa.

W teście Kruskala-Wallisa dla badanej cechy X pobieramy próbę dla k grup. Niech $x_i = \{x_{i1}, x_{i2}, \dots, x_{i,n_i}\}$ oznacza realizację dla i -tej grupy oraz $i = 1, 2, \dots, k$, natomiast $F_i(x)$ niech oznacza rozkład dla i -tej grupy. Na poziomie istotności $0 < \alpha < 1$ tworzymy hipotezę roboczą: $H_0: F_1(x) = F_2(x) = \dots = F_k(x)$ - rozkłady w grupach są identyczne lub różnią się nieistotnie, (brak różnic pomiędzy efektami), wobec hipotezy alternatywnej:

H_1 - istnieją takie i, j że $F_i(x) \neq F_j(x)$ - rozkłady w grupach istotnie się różnią.

Aby zweryfikować te hipotezy najpierw dokonujemy rangowania dla całej próby $\{x_{11}, x_{12}, \dots, x_{1,n_1}, x_{21}, x_{22}, \dots, x_{2,n_2}, x_{31}, x_{32}, \dots, x_{k,n_k}\}$. R_{ij} oznacza rangę dla elementu x_{ij} , $1 \leq i \leq k, 1 \leq j \leq n_i$. Każdej grupie X_i odpowiada ciąg rang $R_i = \{R_{i1}, R_{i2}, \dots, R_{i,n_i}\}$, dla którego wyznaczamy średnią rangę dla i -tej grupy (39):

$$\bar{R}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} R_{ij}. \quad (39)$$

Statystyka testowa jest dana wzorem (40):

$$T = \frac{12}{n(n+1)} \sum_{i=1}^k n_i \left(\bar{R}_i - \frac{n+1}{2} \right)^2, \quad (40)$$

gdzie:

i jest miarą odstępstwa średnich próbkowych rang od $\frac{n+1}{2}$.

Statystyka testowa T ma rozkład χ^2 o $(k-1)$ stopniach swobody. Zbiór krytyczny jest postaci (41):

$$W = [\chi_{1-\alpha}^2(k-1), +\infty), \quad (41)$$

Gdzie:

$\chi_{1-\alpha}^2(k-1)$ jest kwantylem rzędu,

$1 - \alpha$ dla rozkładu χ^2 o $k-1$ stopniach swobody.

Jeżeli $T \in W$ to na poziomie istotności α odrzucamy H_0 na korzyść hipotezy alternatywnej H_1 , zatem rozkłady w grupach różnią się istotnie. Jeżeli $T \notin W$ to na poziomie istotności α nie ma podstaw do odrzucenia H_0 , zatem możemy przyjąć, że rozkłady w grupach są identyczne bądź różnice rozkładów są nieistotne.

Następnie, w przypadku stwierdzenia istotności różnic w grupach, należy dokonać porównania istotności różnic indywidualnie pomiędzy wyszczególnionymi skupieniami. Powyższe wykonywano w oparciu o test post hoc HSD Tukey'a z poprawką Kramera (ang. Tukey's honest significant difference test) [111]. Zastosowanie poprawki umożliwia na porównanie grup o różnej liczności.

Test ten umożliwia wielokrotne porównanie różnic pomiędzy średnimi w wytypowanych grupach. Bazuje on na rozkładzie zbliżonym do rozkładu t-Studenta. W ramach testu zakłada się następującą hipotezę zerową:

$$H_0: \mu_i = \mu_j,$$

wobec hipotezy alternatywnej:

$$H_1: \mu_i \neq \mu_j.$$

Statystyka testowa dana jest wzorem (42):

$$q = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{\frac{MSE}{2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}}, \quad (42)$$

gdzie:

$\bar{x}_i, \bar{x}_j$ – wartości średniej w poszczególnych grupach,

n_i, n_j – liczności w poszczególnych grupach,

MSE – średni kwadrat błędu z analizy wariancji,

Wartości krytyczna q_{kr} dla tego rozkładu przedstawiona jest w tabeli studentyzowanego przy przyjętym poziomie istotności α oraz przyjętych stopniach swobody $N - k, k$, gdzie N – liczebność całkowita, k – liczba średnich [13].

Jeżeli $q > q_{kr}$ to na poziomie istotności α odrzucamy H_0 na korzyść hipotezy alternatywnej H_1 , zatem rozkłady w grupach różnią się istotnie. Jeżeli $q \leq q_{kr}$ to na poziomie istotności α nie ma podstaw do odrzucenia H_0 , zatem możemy przyjąć, że rozkłady w grupach są identyczne bądź różnice rozkładów są nieistotne.

4.2.2. Modelowanie za pomocą metody regresji wielorakiej

W drugim etapie badań przystąpiono do modelowania emisji CO₂ z zastosowaniem metody regresji wielorakiej. Stanowi ona rozwinięcie jednego z powszechnie stosowanych narzędzi statystycznych m.in. w analizie danych oraz procesach modelowania tj. regresji liniowej. Umożliwia ona określenie zależności pomiędzy zmienną zależną (objaśnianą), a jedną lub wieloma zmiennymi niezależnymi (przewidywanymi lub objaśniającymi). Pozwala ona

ustalić w jaki sposób zmiana zmiennych niezależnych wpływa na zmienną zależną. Głównym zadaniem regresji liniowej jest odnalezienie zależności matematycznej, która możliwie jak najdokładniej opisuje relację między badanymi zmiennymi i minimalizuje błędy pomiędzy wartościami przewidywanymi i obserwowanymi.

Modelowanie z wykorzystaniem regresji liniowej jest powszechnie stosowane ze względu na możliwość [112]:

- testowania różnych założeń badawczych oraz utworzenia modelu, który najwierniej opisuje związek pomiędzy badanymi zmiennymi,
- zrozumienia i interpretacji badanego zjawiska oraz występujących w nim zależności,
- prognozowania analizowanych zjawisk.

Przed przystąpieniem do modelowania dostępny zbiór danych podzielono na dwa podzbiory (uczący i testowy). Zbiór uczący wykorzystywany jest do budowy modelu, natomiast zbiór testowy do przeprowadzenia oceny jego dopasowania (obliczenia błędów prognozy). Wyznaczanie modelu regresji liniowej składa się z 4 zasadniczych etapów, tj. specyfikacji, estymacji, weryfikacji, zastosowania. Pierwszym etapem jest specyfikacja, czyli przedstawienie modelu w postaci, która umożliwi jego dalszą analizę i weryfikację empiryczną. Kolejny etap to estymacja, czyli wyznaczenie najlepszych wartości parametrów modelu. Trzeci etap to weryfikacja, czyli sprawdzenie za pomocą metod statystycznych czy otrzymany model jest poprawny (ocena istotności modelu i jego parametrów). Końcowy etap to zastosowanie, poprawny model można wykorzystać, np. do prognozowania [112].

Regresja przyjmuje postać równania lub układu równań. Strukturę modelu regresji przedstawia następująca zależność (43) [112]:

$$Y = F(X_1, X_2, \dots, X_k, \varepsilon) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon, \quad (43)$$

gdzie:

Y – zmienna zależna,

$X_1, X_2, \dots, X_k$ – zmienne niezależne,

ε – składnik losowy,

F – nieznana postać funkcyjna modelu.

Funkcja regresji najczęściej ma postać liniową (regresja liniowa), ale może przyjąć również postać funkcji wykładniczej, potęgowej, logarytmicznej itd. (regresja nieliniowa) [112]. W sytuacji, kiedy analizowana jest zależność pomiędzy jedną zmienną zależną a wieloma zmiennymi niezależnymi mamy do czynienia z regresją wieloraką [112].

Parametry $\beta_0, \beta_1, \beta_k$ modelu (43) nie są znane. Należy je oszacować na podstawie próby badawczej, za pomocą metody najmniejszych kwadratów (MNK). MNK polega na doborze takich parametrów, dla których suma kwadratów odchyleń od modelu jest jak najmniejsza, co opisuje równanie (44) [40], [112]:

$$\sum_{i=1}^k (y_i - \hat{y}_i)^2 = \sum_{i=1}^k (y_i - \beta_1 x_i - \beta_0)^2 \rightarrow \min, \quad (44)$$

gdzie:

y_i – i -ta wartość empiryczna zmiennej zależnej Y ,

x_i – i -ta wartość zmiennej niezależnej X ,

$\hat{y}_i$ – i -ta wartość teoretyczna wyznaczona na podstawie modelu,

β_0 – wyraz wolny równania regresji,

$\beta_1, \beta_2, \dots, \beta_k$ – nieznane wartości parametrów modelu regresji.

Oszacowana funkcja regresji przyjmuje wówczas postać (45) [40], [112]:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k, \quad (45)$$

gdzie:

$x_1, x_2, \dots, x_k$ – zmienne objaśniające (niezależne),

b_j – parametry modelu (współczynniki regresji) dla j -tej zmiennej,

$j = 1, 2, \dots, k$,

ε – składnik losowy.

Wartość estymatora j -tego współczynnika regresji liniowej b_j umożliwia określenie o ile jednostek wzrośnie (lub zmaleje) wartość zmiennej zależnej, gdy wartość j -tej zmiennej niezależnej wzrośnie o 1 jednostkę, przy założeniu, że pozostałe zmienne w modelu nie zmieniają swojej wartości. Własność ta ma duże znaczenie praktyczne z uwagi na nieskomplikowaną i bezpośrednią interpretację modelu.

Model regresji wielorakiej wymaga spełnienia następujących założeń [112]:

1. Liniowa zależność zmiennej niezależnej względem parametrów modelu regresji.
2. Liczba obserwacji n musi być większa od liczby szacowanych parametrów, tj. $n \geq k + 1$.
3. Poszczególne zmienne niezależne nie są kombinacjami liniowymi pozostałych zmiennych niezależnych.
4. Składnik losowy ma wartość oczekiwaną równą 0 dla wszystkich $i = 1, 2, \dots, n$.
5. Wariancja składnika losowego jest identyczna dla wszystkich obserwacji (homoskedastyczność).

6. Składniki losowe są nieskorelowane.
7. Składnik losowy cechuje się rozkładem losowym dla dowolnych obserwacji.

Weryfikacji modelu dokonuje się przez [116]:

1. Sprawdzenie istotności modelu regresji wielorakiej

W celu weryfikacji istotności modelu regresji wielorakiej wykorzystuje się uogólniony test Walda. W teście tym zakłada się hipotezę zerową:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0,$$

wobec alternatywnej:

$$H_1: \beta_1 \neq \beta_2 \neq \dots \neq \beta_k \neq 0.$$

Wartość statystyki testowej obliczana jest według wzoru (46) [111]:

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - (k + 1)}{k}, \quad (46)$$

gdzie:

R^2 – współczynnik determinacji modelu,

n – liczba obserwacji,

k – liczba parametrów modelu.

Statystyka testowa F ma rozkład Fishera-Snedecora o $(k - 1)$ i $n - (k + 1)$ stopniach swobody. Wartość krytyczna przyjmuje wartość (47):

$$F^* = F_{\alpha, (k-1), n-(k+1)}. \quad (47)$$

Obszar krytyczny jest postaci (48):

$$W = [F_{\alpha}(k - 1, n - (k + 1)), +\infty). \quad (48)$$

Jeżeli $F \in W$ to na poziomie istotności należy odrzucić hipotezę zerową H_0 na korzyść H_1 , co oznacza, że istnieje taka zmienna objaśniająca X_i , która ma statystycznie istotny wpływ na zmienną objaśnianą Y . Jeżeli $F \notin W$ to nie ma podstaw do odrzucenia hipotezy zerowej H_0 , co oznacza, że nie ma takiej zmiennej X_i , która ma statystycznie istotny wpływ na zmienną objaśnianą Y , a zatem wszystkie zmienne są nieistotne dla rozpatrywanego modelu.

2. Sprawdzenie istotności parametrów modelu

Do sprawdzenia istotności parametrów modelu zastosowano test t-Studenta. W ramach testu sformułowano hipotezę zerową:

$H_0: \beta_i = 0$, brak statystycznie istotnego wpływu zmiennej objaśniającej X_i na zmienną Y ,
oraz alternatywną

$H_1: \beta_i \neq 0$, występuje statystycznie istotny wpływ zmiennej objaśniającej X_i na zmienną Y ,
które zweryfikowano dla przyjętego poziomu istotności α .

Wartość statystyki testowej obliczana jest według wzoru (49) [111]:

$$t = \frac{|\beta_i|}{SE}, \quad (49)$$

gdzie:

$$SE = \frac{S_{yx}}{sd_x \sqrt{n-1}}, \quad (50)$$

$$s_{yx} = sd_y \sqrt{\frac{n-1}{n-2} (1-R^2)}, \quad (51)$$

sd_x , sd_y – odchylenie standardowe wartości zmiennych odpowiednio X oraz Y .

Statystyka testowa ma rozkład t-Studenta o $(n-2)$ stopniach swobody. Wartość krytyczna przyjmuje wartość (52):

$$t^* = t_{\alpha, n-2}. \quad (52)$$

Obszar krytyczny jest postaci (53):

$$W = [t_{\alpha}(n-2), +\infty). \quad (53)$$

Jeżeli $F \in W$ to na poziomie istotności należy odrzucić hipotezę zerową H_0 na korzyść H_1 , co oznacza, że zmienna objaśniająca X_i (z odpowiadającym jej parametrem β_i) ma statystycznie istotny wpływ na zmienną objaśnianą Y . Jeżeli $F \notin W$ to nie ma podstaw do odrzucenia hipotezy zerowej H_0 , co oznacza, że zmienna X_i ma statystycznie nieistotny wpływ na zmienną objaśnianą Y .

3. Brak współliniowości (nadmiarowości) zmiennych niezależnych

Współliniowość zmiennych dotyczy sytuacji, w której zmienne są ze sobą silnie skorelowane. W przypadku liniowego przebiegu zmiennych analizę korelacji przeprowadza się w oparciu o współczynnik korelacji Pearsona. Współczynnik ten liczony jest według poniższego wzoru (54):

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (54)$$

gdzie:

x_i oraz y_i oznaczają wartości empiryczne poszczególnych zmiennych (obserwacje, pomiary), natomiast $\bar{x}$ oraz $\bar{y}$ oznaczają wartości średnie poszczególnych zmiennych.

Interpretację siły współczynnika korelacji przedstawiono w tabeli 11.

Tabela 11. Interpretacja współczynnika korelacji Pearsona [116]

Współczynnik korelacji r	Siła związku korelacyjnego
0,0 – 0,2	Brak
0,2 – 0,4	Słaba
0,4 – 0,7	Średnia
0,7 – 0,9	Silna
0,9 – 1,0	Bardzo silna

Znak współczynnika korelacji mówi o kierunku związku pomiędzy zmiennymi. Gdy jest dodatni oznacza, że wzrost (spadek) wartości jednej cechy powoduje wzrost (spadek) wartości drugiej (relacja wprost proporcjonalna). Natomiast, kierunek ujemny oznacza, że wraz ze wzrostem (spadkiem) wartości cechy występuje spadek (wzrost) wartości drugiej (relacja odwrotnie proporcjonalna).

4. Diagnostyka reszt modelu:

- Badanie zgodności rozkładu składnika losowego z rozkładem normalnym:

Badanie zgodności rozkładów wykonano z zastosowaniem testu chi-kwadrat zgodności, który służy do porównania rozkładu empirycznego zmiennej losowej dyskretnej z rozkładem teoretycznym [13]. Niech $X: \Omega \rightarrow S = \{S_1, S_2, \dots, S_k\}$:

$$P(X = S_i) = p_i, \quad (55)$$

dla $p_i \geq 0$ oraz $\sum_{i=1}^k p_i = 1$. Dla realizacji zmiennej losowej (obserwowanej cechy) $\{x_1, x_2, \dots, x_n\}$ wyznaczamy liczebność każdej klasy, gdzie $x_i \in S$, a $i = 1, 2, \dots, n$. Niech n_i oznacza liczebność klasy dla realizacji S_i oraz $\sum_{i=1}^k n_i = n$. Na poziomie istotności $0 < \alpha < 1$ tworzymy hipotezę roboczą:

$H_0: P(X = S_i) = p_i$ dla $i = 1, 2, \dots, k$ - badana cecha X ma rozkład dyskretny $(p_1, p_2, \dots, p_k)$, wobec hipotezy alternatywnej:

H_1 : istnieje taka klasa j dla której $P(X = S_i) \neq p_i$.

Statystyka testowa (56) [13]:

$$\chi^2 = \sum_{j=1}^k \frac{(n_i - np_i)^2}{np_i}, \quad (56)$$

ma rozkład chi-kwadrat o $(k - 1)$ stopniach swobody. Obszar krytyczny jest postaci (57):

$$W = [\chi_{1-\alpha}^2(k - 1), +\infty), \quad (57)$$

gdzie:

$\chi_{1-\alpha}^2(k - 1)$ oznacza kwantyl rzędu $1 - \alpha$ dla rozkładu χ^2 o $k - 1$ stopniach swobody.

Jeżeli $\chi^2 \in W$ to na poziomie istotności α odrzucamy H_0 na korzyść hipotezy alternatywnej (rozkład badanej cechy X istotnie różni się od rozkładu empirycznego). Jeżeli $\chi^2 \notin W$ to na poziomie istotności α nie ma podstaw do odrzucenia H_0 , możemy przyjąć, że cecha X ma rozkład $P(X = S_i) = p_i, i = 1, 2, \dots, k$.

- Jednorodność wariancji składnika losowego dla kilku wariantów zmiennych objaśniających (homoscedastyczność):

Zjawisko heteroskedastyczności w modelu ma miejsce, kiedy wariancja składnika losowego nie jest taka sama dla wszystkich obserwacji. Do weryfikacji powyższego zastosowano test White'a, w ramach którego przyjęto:

hipotezę zerową [111]:

$$H_0: \sigma_i^2 = \sigma^2,$$

oraz hipotezę alternatywną:

$$H_1: \sigma_i^2 \neq \sigma^2.$$

Statystyka testowa (58) [111]:

$$W = nR^2, \quad (58)$$

ma asymptotyczny rozkład χ^2 z liczbą swobody k .

Obszar krytyczny jest postaci (59):

$$W = [\chi_\alpha^2(k), +\infty), \quad (59)$$

gdzie:

$\chi_\alpha^2(k)$ oznacza kwantyl rzędu $1 - \alpha$ dla rozkładu χ^2 o k stopniach swobody.

Jeżeli $\chi^2 \in W$ to na poziomie istotności α odrzucamy H_0 na korzyść hipotezy alternatywnej (brak jest jednorodności wariancji składnika losowego). Jeżeli $\chi^2 \notin W$ to na poziomie istotności α nie ma podstaw do odrzucenia H_0 , możemy przyjąć, że wariancja składnika losowego jest jednorodna.

- Brak autokorelacji składnika losowego (test Durбина-Watsona):

W przypadku wystąpienia autokorelacji składników losowych, estymatory przestają być najefektywniejszymi estymatorami. Badanie występowania tego zjawiska dokonuje się przez weryfikację hipotezy za pomocą testu Durбина-Watsona [112].

Test Durбина-Watsona (DW) [13] służy do weryfikacji hipotezy o braku autokorelacji rzędu pierwszego w szeregu reszt $\{\varepsilon_t\}_{t \in N}$. Na poziomie istotności α formułujemy hipotezę zerową.

$H_0: \rho = 0$ – współczynnik autokorelacji reszt modelu jest statystycznie równy zero,

wobec hipotezy alternatywnej:

$H_0: \rho \neq 0$ – współczynnik autokorelacji elementów szeregu jest statystycznie różny od zera. Dla ciągu realizacji wyznaczamy zależność korelacyjną $\varepsilon_t = \rho\varepsilon_{t-1} + \epsilon_t$ pomiędzy elementami ε_t i ε_{t-1} , która wyraża się wzorem (60):

$$\rho = \frac{\sum_{t=1}^{n-1} \varepsilon_t \varepsilon_{t+1}}{\sum_{t=1}^n \varepsilon_t^2}. \quad (60)$$

Statystyka testowa Durbina-Watsona jest postaci (61):

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}. \quad (61)$$

Statystyka DW przyjmuje wartości od 0 (dodatnia autokorelacja liniowa) do 4 (ujemna autokorelacja liniowa). Hipoteza zerowa zostaje zachowana, jeśli statystyka ma wartość bliską 2 (brak autokorelacji liniowej). Z tablic testu Durbina-Watsona otrzymujemy wartości krytyczne dolną d_L i górną d_U . Dla $\rho > 0$ jeżeli $d > d_U$ to na poziomie istotności α nie ma podstaw do odrzucenia H_0 , jeżeli $d < d_L$ to H_0 odrzucamy na korzyść H_1 . Jeżeli $d_L \leq d \leq d_U$, to za pomocą testu Durbina-Watsona nie można zweryfikować istotności korelacji elementów szeregu dla przesunięcia pierwszego i należy zastosować inny test.

Dla $\rho < 0$ jeżeli $4 - d > d_U$ to na poziomie istotności α nie ma podstaw do odrzucenia H_0 , jeżeli $4 - d < d_L$ to H_0 odrzucamy na korzyść H_1 . Jeżeli $d_L \leq 4 - d \leq d_U$ to za pomocą testu Durbina-Watsona nie można zweryfikować istotności korelacji elementów szeregu dla przesunięcia pierwszego i również należy zastosować inny test.

W ostatnim etapie opracowania modelu należy dokonać jego oceny. Do najważniejszych miar w tym zakresie należą [112], [40]:

- współczynnik determinacji R^2 (62):

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (62)$$

gdzie:

n – liczba obserwacji,

y_i – wartość empiryczna zmiennej Y ,

$\hat{y}_i$ – wartość teoretyczna zmiennej Y wyznaczona na podstawie modelu,

$\bar{y}$ – średnia arytmetyczna wartości zmiennej Y .

- Średnia kwadratów reszt (MSE) (63):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (63)$$

- Średni błąd bezwzględny (MAE) (64):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (64)$$

- Względne odchylenie przeciętne (MRAE) (66):

$$r_i = \frac{|y_i - \hat{y}_i|}{y_i - f_t^*}, \quad (65)$$

$$MRAE = \frac{1}{n} \sum_{i=1}^n |r_i|, \quad (66)$$

gdzie:

f_t^* - wartość prognozy otrzymanej z modelu referencyjnego (najczęściej przy stosowaniu powyższej miary jako model referencyjny wykorzystuje się model naiwny).

- Średni bezwzględny błąd procentowy (MAPE) (67):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\%. \quad (67)$$

- Kryterium informacyjne AIC (68):

$$AIC = -2 \ln L + 2k, \quad (68)$$

gdzie:

k - liczba parametrów w modelu,

L – funkcja wiarygodności.

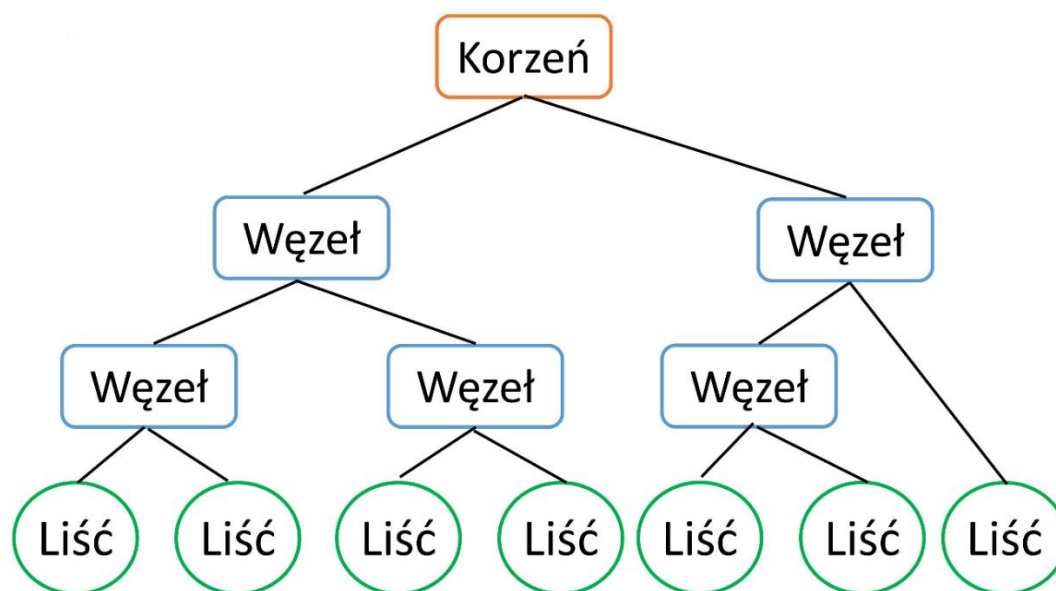
Wskaźniki AIC to kompromis pomiędzy dopasowaniem modelu a liczbą parametrów. Do predykcji wybieramy model, dla którego wartość wskaźnika jest najmniejsza.

4.2.3. Modelowanie z wykorzystaniem metody drzew decyzyjnych

Alternatywą dla klasycznych modeli analitycznych są metody uczenia maszynowego. W odróżnieniu od tradycyjnych metod analitycznych, które opierają się na predefiniowanych regułach, statystyce opisowej i dedukcyjnych modelach matematycznych, uczenie maszynowe pozwala na automatyczne uczenie i poprawę wydajności na podstawie zgromadzonych danych. Popularną metodą analizy danych z wykorzystaniem technik uczenia maszynowego są drzewa decyzyjne, które umożliwiają podejmowanie decyzji na podstawie znajomości pewnego zbioru danych wejściowych, tzw. atrybutów. Celem stosowania tej metody jest graficzne określenie planu rozwiązania problemu w postaci tzw. drzewa [82].

Struktura drzewa decyzyjnego składa się z wierzchołków (punkt górny, z którego następują odgałęzienia), węzłów oraz liści (rys. 34). Wierzchołek na najwyższym poziomie nazywany jest korzeniem, natomiast liście stanowią jego końcową część, której przypisane są

odpowiednie klasy. Z wierzchołków wychodzą odgałęzienia, które łączą się z węzłami stanowiącymi łączniki pomiędzy odgałęzieniami na różnych poziomach [82]. Omawiana konstrukcja posiada strukturę drzewiastą, w której węzły wewnętrzne zawierają tzw. testy na wartościach atrybutów i z każdego z nich wychodzi tyle gałęzi, ile jest możliwych wyników testu w tym węźle. Z kolei liście zawierają decyzje o klasie obiektów. Celem drzewa decyzyjnego jest segregacja wspomnianych obiektów według klas [71], [31], [82].



Rys. 34. Schemat budowy drzewa decyzyjnego [39]

W celu omówienia zasady działania drzew decyzyjnych należy opisać szereg etapów związanych z ich budową. W pierwszym etapie następuje utworzenie modelu drzewiastego na podstawie dostępnych przykładów (pojedyncze wiersze danych), który jest następnie używany do klasyfikacji. W wymienionym procesie dokonuje się porównania wartości danego atrybutu z określonym progiem. Wynik tego działania jest przekazywany do kolejnego węzła niższego poziomu, natomiast sama operacja jest powtarzana aż do momentu, gdy wszystkie przypadki zostaną sklasyfikowane poprawnie. Z kolei w trybie odtwarzania w celu klasyfikacji danego obiektu wartości jego atrybutu są testowane zgodnie z informacją zawartą w drzewie decyzyjnym, natomiast w procesie testowania pokonywana jest występująca w drzewie ścieżka od wierzchołka do korzenia, oznaczająca klasę, do której zalicza się określony obiekt. W kolejnym kroku dokonuje się oceny dokładności modelu klasyfikatora na zbiorze danych testowych. W tym przypadku dokładność modelu określa się przez procentową liczbę trafnych (poprawnych) klasyfikacji (prawidłowych porównań etykiety klasy z etykietą określoną przez model). Po wykonaniu oceny wyników, jeśli dokładność modelu zostanie uznana za wystarczającą, model może zostać użyty w celu klasyfikacji dla kolejnych przypadków.

W przeciwnym razie następuje dalsze przetwarzanie modelu, najczęściej polegające na redukcji drzewa, tzw. przycinania aż do uzyskania przyjętych wyników testowania [71], [82].

Podczas budowy drzewa decyzyjnego z M atrybutów wejściowych, które wchodzą w skład wektora y należy wybrać wyłącznie jeden, np. y_i , dla którego i stanowi wartość progu podziału zbioru danych p_i . W tym przypadku jeden z analizowanych obszarów zawiera dane, które spełniają warunek $y_i < p_i$, natomiast drugi $y_i > p_i$. Decyzja dotycząca wyboru wartości progu klasyfikacyjnego powinna utworzyć obszary możliwie jednorodne klasowo, jednak w praktyce jest to trudne do osiągnięcia, dlatego dobór progu realizuje się minimalizując miarę rozkładu klas decyzyjnych w obu wymienionych obszarach [82].

Do wymienionych miar zalicza się m.in. współczynnik Giniego (69), który definiuje się dla obszaru A w następujący sposób [82]:

$$I(A) = 1 - \sum_{j=1}^N r_j^2, \quad (69)$$

gdzie:

N – liczba klas,

r_j – prawdopodobieństwo wystąpienia j -tej klasy w obszarze A .

Do wskaźników rozkładu klas decyzyjnych zalicza się również entropia (70), którą można opisać dla obszaru A przy pomocy poniższej zależności [82]:

$$E(A) = - \sum_{j=1}^N r_j \log_2 r_j, \quad (70)$$

gdzie:

N – liczba klas,

r_j – prawdopodobieństwo wystąpienia j -tej klasy w obszarze A .

Jeśli rozpatrywany obszar A zawiera jedynie dane należące do jednej klasy, to w takim przypadku obydwie miary są równe zero. W sytuacji, gdy należy podzielić obszar A na dwa podzbiory A_1 i A_2 , to łączna miara opisuje podział stanowiący sumę wagową dwóch podobszarów. W przypadku miary Giniego (71) omawianą pełną miarę można zapisać w następujący sposób [82]:

$$I(A) = \frac{l_{A1}}{l_A} I(A_1) + \frac{l_{A2}}{l_A} I(A_2), \quad (71)$$

gdzie:

l_{A1} , l_{A2} - liczba elementów zawartych w obszarze A_1 i A_2 ,

l – całkowita liczba elementów w obu obszarach.

W kontekście miary entropii (72) postępuje się analogicznie jak w przypadku miary Giniego [82]:

$$E(A) = \frac{l_{A1}}{l_A} E(A_1) + \frac{l_{A2}}{l_A} E(A_2), \quad (72)$$

gdzie:

l_{A1}, l_{A2} - liczba elementów zawartych w obszarze A_1 i A_2 ,

l – całkowita liczba elementów w obu obszarach.

Algorytm tworzenia drzewa decyzyjnego można opisać słownie w postaci następujących kroków [82]:

1. Należy opracować zbiór danych uczących w postaci określonej liczby n par uczących (x, d) , gdzie x oznacza wektor atrybutów wejściowych (x_i dla $i = 1, 2, \dots, m$), a d oznacza etykietę klasy. Z poszczególnych wektorów x jest tworzona macierz X o wymiarach $n \times m$. Następnie należy posortować elementy kolumn w wymienionej macierzy w porządku rosnącym oraz zachować pierwotną macierz w oryginalnym porządku elementów. W kolejnym kroku dla wszystkich uporządkowanych kolumn określa się średnie arytmetyczne dla dwóch sąsiednich wartości elementów tworzące wstępne wartości progów liczbowych. Z wszystkich tych wartości ostatecznie tworzy się macierz potencjalnych progów P .
2. Wartość progowa testu dla poszczególnych zmiennych jest określana na podstawie macierzy P i stanowi wartość minimalnej miary Giniego lub entropijnej obliczanej dla danego obszaru danych. W rozpatrywanym punkcie drzewa tworzy się węzeł dla zmiennej x_i przyjmujący minimalną wartość przyjętej miary z określonym dla niej progiem.
3. Dla rozpatrywanego węzła poszczególne przykłady dzieli się na 2 podmacierze w wyniku porównania zmiennej x_i z przyjętym progiem.
4. Następnie etapy są realizowane dla konkretnej podmacierzy przykładów. W przypadku, w którym przykłady zaliczają się do jednej klasy następuje koniec podziału, tzn. węzeł staje się liściem drzewa z oznaczeniem danej klasy, natomiast algorytm realizuje kolejne operacje na najbliższej podmacierzy (nieposortowanej).
5. W przypadku, gdy w wyniku zastosowania całkowitej procedury niektóre przykłady nie mogą być przydzielone jednoznacznie do danych klas, to tworzy się tzw. liść mieszany, który zawiera kategorię większościową.

Ważnym parametrem związanym z tworzeniem drzew decyzyjnych jest kryterium stopu, którego spełnienie powoduje zatrzymanie procesu konstrukcji drzewa decyzyjnego w przypadku, gdy zbiór obiektów: jest pusty, składa się z elementów należących tylko do jednej klasy decyzyjnej oraz zbiór możliwych do użycia testów jest pusty [31], [82].

Po utworzeniu drzewa decyzyjnego należy ocenić jego jakość, czego można dokonać na kilka sposobów. Jednym z nich jest ocena rozmiaru utworzonego drzewa. W tym przypadku analizuje się parametry, tj. liczba węzłów, wysokość drzewa, liczbę liści itd. Dąży się do minimalizacji wymienionych wartości. Kolejnym sposobem jest ocena dokładności klasyfikacji na zbiorze treningowym. W omawianym przypadku rozpatrywaną miarą jest stosunek poprawnie sklasyfikowanych obserwacji uczących w stosunku do wszystkich obserwacji biorących udział w uczeniu. Dodatkowo bierze się także pod uwagę w analogiczny sposób dokładność klasyfikacji na zbiorze testowym, przy czym w tym przypadku dane te nie występują w danych uczących [31], [82].

Proces tworzenia drzew decyzyjnych może podlegać optymalizacji pod kątem wybranych parametrów, natomiast sam proces poszukiwania optymalnej struktury jest trudny do realizacji w czasie wielomianowym (problem typu NP). Z wymienionego powodu wynika potrzeba tworzenia dedykowanych algorytmów heurystycznych. Jedną z technik mających za zadanie poprawić zdolność generalizacji, a więc zmniejszyć błędy klasyfikacji na podstawie danych nie biorących udziału w procesie uczenia jest tzw. przycinanie. Najczęściej mniejsze drzewa uzyskują mniejszą dokładność odwzorowania danych pochodzących ze zbioru uczącego, co może prowadzić do większej dokładności klasyfikacji dla danych pochodzących ze zbioru testowego. Stosuje się w tym przypadku technikę upraszczania struktury drzewa, która polega na zastąpieniu fragmentu drzewa pojedynczym liściem z najczęściej występującą etykietą w tym fragmencie lub zmniejszeniu złożoności fragmentu drzewa [31], [82].

Drzewa decyzyjne powstały na potrzeby rozwiązywania zadania klasyfikacji, jednak można ją również stosować w odniesieniu do zadania regresji. Dokonuje się tego jako rozszerzenie możliwości drzew decyzyjnych, które polega na kwantyzacji poziomów wartości w ramach zadanego zakresu. Sprowadza to tym samym omawiane zadanie regresji do zadania przypisania wartości poszukiwanej do określonego poziomu kwantyzacji (analogiczne do klasyfikacji) [82].

Przedstawiona teoria dotycząca drzew decyzyjnych opisuje ich strukturę, proces tworzenia oraz zasadę działania, jednak w celu efektywnego wykorzystania potencjału drzew losowych do modelowania złożonych procesów stosuje się wybrane algorytmy uczenia maszynowego, m.in. algorytm CART (z ang. Classification and Regression Trees), lasy losowe

(z ang. Random Forests), czy drzewa wzmacniane (z ang. Boosted Decision Tree). Wszystkie wcześniej wymienione metody opierają się na koncepcji hierarchicznego podziału przestrzeni cech. Metoda CART tworzy binarne drzewa decyzyjne przez rekurencyjny podział danych na coraz bardziej jednorodne podzbiory. Lasy losowe rozszerzają tę koncepcję, tworząc wiele niezależnych drzew na podstawie losowo wybranych podzbiorów danych i cech. Z kolei drzewa wzmacniane, tworzą iteracyjnie sekwencję drzew, w której każde kolejne drzewo koncentruje się na poprawie błędów poprzednich. Każda z metod wykorzystuje strukturę drzewiastą, różnią się jednak metodą konstrukcji modelu końcowego [71], [31], [82]. W kolejnej części rozdziału wymienione metody zostaną omówione szczegółowo.

Algorytm CART

Pierwszą z wymienionych metod jest algorytm CART, który służy m.in. do konstrukcji binarnych drzew klasyfikacyjnych, tzn. takich dla których każdy węzeł dzieli się na dwa węzły potomne. Podstawą działania algorytmu CART jest to, iż podział danych w węzłach zależy wyłącznie od jednej zmiennej decyzyjnej, natomiast jego zatrzymanie następuje, gdy odpowiedź na dany test (pytanie) nie tworzy kolejnego podziału i nie wyznacza kolejnych testów. Charakterystyczne dla tej metody jest również to, iż po utworzeniu podziałów, w każdej warstwie jest określany błąd klasyfikacji. Zaletą metody CART jest łatwość interpretacji danych, wynikająca z jej binarnego charakteru oraz możliwość stosowania na niepełnych zbiorach danych [90].

Metoda CART wykorzystuje miarę Giniego jako metodę podziału. W celu kontroli procesu zatrzymania procesu podziału stosuje się najczęściej parametr minimalnej liczebności węzłów. W rozpatrywanym przypadku proces jest powtarzany tak długo, aż węzły końcowe stają się jednorodne lub zawierają określoną wcześniej liczbę przypadków. W przypadku rozmiaru struktury, dąży się do tego, aby drzewo było możliwie najprostsze, przy czym powinno zachowywać złożoność opisywanych danych. W związku z tym w pierwszej kolejności tworzony jest model o maksymalnej liczbie liści, którego określone fragmenty następnie są przycinane, minimalizując przy tym wartość błędu klasyfikacji. Wspomniane przycinanie umożliwia porównanie modelu pełnego oraz zredukowanego i wybór takiej wersji modelu zredukowanego, dla którego różnica błędów klasyfikacji dla obydwu wymienionych modeli jest najmniejsza [90].

Podczas budowy modelu CART zdefiniowano ustawienia brzegowe prezentowanego rozwiązania:

- reguła stopu – przytnij według wariancji,

- minimalna liczebność węzła końcowego $n = 100$ (w trybie ruszania pojazdu) lub $n = 300$ (w trybach jazdy miejskiej i jazdy pozamiejskiej),
- maksymalna liczba poziomów (głębokość) drzewa $n = 10$,
- maksymalna liczba węzłów $n = 1000$,
- szacowanie błędu za pomocą 10-krotnej walidacji testowej.

Drzewa wzmacniane

Główną koncepcją działania drzew wzmacnianych jest tworzenie serii prostych binarnych drzew decyzyjnych (jeden korzeń i dwóch potomków), w przypadku których każde następne drzewo służy do predykcji reszt generowanych przez poprzednie. W kolejnych etapach procesu wyznacza się najlepszy podział danych i określa odchyłki wartości aktualnie obserwowanych w stosunku do wartości średnich (średnia arytmetyczna reszty dla każdego podziału). Algorytm stosuje metodę stochastycznego wzmacniania gradientowego, która zakłada, że każde kolejne drzewo jest tworzone na podstawie losowej próby zawierającej 50% całego zbioru danych. Wprowadzenie wspomnianej losowości służy zabezpieczeniu metody przeciwko przeuczeniu i tworzy tym samym modele z dużą zdolnością generalizacji i własnościami predykcyjnymi. Wobec tego algorytm również umożliwia uzyskanie dopasowania pomiędzy wartościami obserwowanymi i przewidywanymi, szczególnie jeśli zależność pomiędzy nimi jest złożona (np. nieliniowa) [32], [52], [53].

Podczas budowy modelu drzew wzmacnianych zdefiniowano ustawienia brzegowe prezentowanego rozwiązania:

- liczba losowanych predyktorów – 3,
- proporcja losowej próby testowej – 100%,
- liczba drzew – 100,
- zatrzymanie uczenia – procentowy spadek błędu 5%,
- minimalna liczność potomka – $n = 100$ (w trybie ruszania pojazdu) lub $n = 300$ (w trybach jazdy miejskiej i jazdy pozamiejskiej),
- maksymalna liczba poziomów (głębokość) drzewa $n = 10$,
- maksymalna liczba węzłów $n = 1000$.

Lasy losowe

Lasy losowe stanowią odmianę grupy drzew decyzyjnych polegającą na implementacji jednocześnie wielu losowo utworzonych drzew decyzyjnych, z których każde, współpracując ze sobą, ostatecznie tworzy odpowiedź modelu. Metoda w celu predykcji wartości zmiennej zależnej wykorzystuje wiele nieskomplikowanych pod względem struktury drzew decyzyjnych. Wymienione drzewa dokonują przewidywania wartości zmiennej zależnej na

podstawie losowego zbioru atrybutów. Proces tworzenia modelu lasów losowych jest sparametryzowany, tzn. użytkownik określa wcześniej parametry, tj. złożoność drzew składowych, maksymalną liczbę drzew tworzących model, maksymalną głębokość drzewa, maksymalną liczbę węzłów, kryterium zakończenia procesu uczenia itd. Z uwagi na zastosowaną losowość w zasadzie działania metody, model lasu losowego jest często stosowany na dużych zbiorach danych, bez konieczności redukcji i usuwania obserwacji [82].

Zakładając, że liczba przykładów uczących (obserwacji) jest równa n , gdzie każda obserwacja ma wymiar N odpowiadający maksymalnej liczbie atrybutów, to metoda lasu losowego na dowolnym poziomie wykorzystuje jedynie $m < N$ losowo wybranych atrybutów. Zbiór uczący danych n_p jest tworzony z losowo wybranych n przykładów z zastępowaniem, co oznacza, iż identyczna próbka może być losowana wielokrotnie, natomiast zbiór testujący-weryfikujący składa się z pozostałych próbek. Proporcje zbiorów uczących/ testujących ustala się przed rozpoczęciem procesu tworzenia lasu losowego [82].

Zasada działania lasu losowego polega na równoległym tworzeniu k drzew decyzyjnych, dla których podział danych na części uczącą i testującą jest wykonywany losowo i niezależnie dla każdego drzewa. W przypadku procesu uczenia dla każdego węzła jedynie dla m losowo wybranych zmiennych podejmuje się decyzję klasyfikacyjną. Podstawą jej podjęcia jest najczęściej liniowa kombinacja atrybutów występujących w określonym węźle decyzyjnym, co przedstawia poniższa zależność (73) [82]:

$$p = \sum_{i=1}^m w_i x_i, \quad (73)$$

gdzie:

w_i – wartość i -tej wagi,

x_i – wartość i -tego atrybutu.

Dobór wag w_i podlega procesowi optymalizacji, który zapewnia maksymalny zysk klasyfikacyjny. Omawiana procedura jest wykonywana we wszystkich poziomach decyzyjnych w pętli, aż do momentu otrzymania najlepszego wyniku klasyfikacji w oparciu o dane dla rozpatrywanego poziomu. Skutkuje to powstaniem gotowych i rozbudowanych drzew, bez konieczności późniejszego przycinania, które same w sobie tworzą las k drzew. W przypadku danych testowych wynik klasyfikacji jest określany przez operację głosowania większościowego dla wszystkich drzew na podstawie danych zwartych w aktualnym wektorze [82].

Wykazano, iż górne oszacowanie błędu generalizacji lasu losowego, przy założeniu powstania dostatecznie dużej liczby drzew, jest możliwe przy pomocy zależności (74) [82]:

$$\varepsilon_g \leq \frac{\rho(1 - s^2)}{s^2}, \quad (74)$$

gdzie:

ρ – uśredniony współczynnik korelacji między drzewami,

s – współczynnik mierzący siłę drzew lasu.

Współczynnik siły drzew lasu s stanowi wartość średniej arytmetycznej probabilistycznej miary dokładności klasyfikacyjnej kolejnych drzew decyzyjnych w rozpatrywanej strukturze lasu, przy założeniu, iż do określenia dokładności i -tego klasyfikatora stosuje się prawdopodobieństwo (75) [82]:

$$s_i = Pr(y(x) = d) - \max Pr(y(x) \neq d), \quad (75)$$

gdzie:

$P(x)$ – prawdopodobieństwo zdarzenia x ,

$y(x)$ – odpowiedź drzewa na wektor x ,

d – oznaczenie klasy, do której należy wektor x .

Z powyższej zależności wynika, iż prawdopodobieństwo podjęcia właściwej decyzji przez drzewo zwiększa się wraz ze wzrostem wartości marginesu. Z kolei górne oszacowanie błędu ε_g jest proporcjonalne w stosunku do wartości korelacji poszczególnych drzew w lesie. Wartość wymienionego błędu również wzrasta, gdy siła drzew lasu zmniejsza się. Z kolei losowy wybór atrybutów w odniesieniu do kolejnych węzłów drzewa zmniejsza korelację dla poszczególnych drzew. Ponadto zwiększenie liczby zmiennych m prowadzi do wzrostu wartości współczynników sił poszczególnych drzew, co również wpływa na ich średnią korelację działania w lesie. Podsumowując, wybór wartości zmiennej m wymaga wykonania szeregu prób w celu uwzględnienia wpływu aspektów korelacji oraz siły poszczególnych drzew [82].

Do głównych zalet metody lasu losowego należą odporność na zjawisko przeuczenia, możliwość oszacowania ważności zmiennych w procesie podejmowania decyzji, duża szybkość uczenia drzew w lesie, większa dokładność działania w stosunku do pojedynczego drzewa, możliwość estymacji brakujących danych oraz pracy na dużych zbiorach danych [82].

Podczas budowy modelu drzew wzmacnianych zdefiniowano ustawienia brzegowe prezentowanego rozwiązania:

- współczynnik uczenia – 0,1;
- liczba drzew – 200;
- proporcja losowej próby testowej – 10%;
- proporcja dla podprób – 50%;
- zatrzymanie uczenia – procentowy spadek błędu 5%;
- minimalna liczność potomka – $n = 100$ (w trybie ruszania pojazdu) lub $n = 300$ (w trybach jazdy miejskiej i jazdy pozamiejskiej);
- minimalne n potomka – 1;
- maksymalna liczba poziomów (głębokość) drzewa $n = 10$;
- maksymalna liczba węzłów $n = 10$.

5. MODELE EMISJI CO₂ DO ŚRODOWISKA

5.1. Wstępna analiza danych oraz przebiegu zmiennej objaśnianej i objaśniającej

Modelowanie emisji CO₂, przeprowadzonej zgodnie z algorytmem przedstawionym w rozdziale czwartym, rozpoczęto od wstępnej analizy danych oraz badania przebiegu zmiennych objaśniających oraz objaśnianej. Na tej podstawie dokonano wyboru odpowiednich modeli matematycznych. W pierwszym kroku obliczono podstawowe statystyki opisowe (12-15) badanych zmiennych (tabela 12).

Tabela 12. Statystyki opisowe wybranych zmiennych

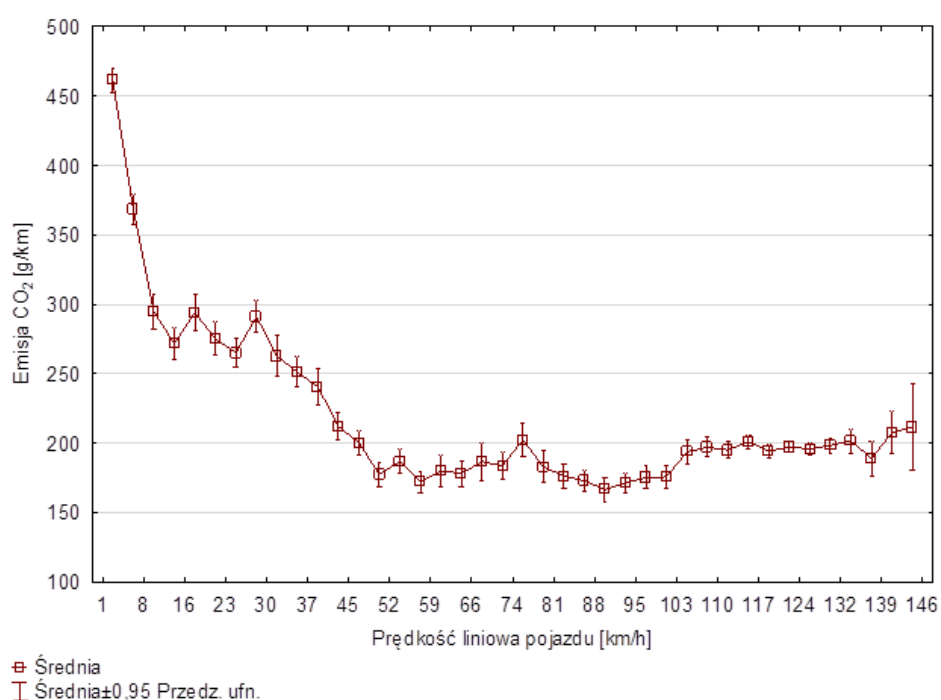
Badana zmienna	Średnia	Mediana	Odch.std	Skośność	Kurtoza
prędkość liniowa pojazdu v [km/h]	67,69	63,00	39,69	0,07	-0,32
względne obciążenie silnika L_% [%]	39,01	36,86	20,91	0,61	-0,22
prędkość obrotowa silnika n [obr/min]	1 839,09	1 865,75	540,79	-0,33	-0,56
moc silnika P [kW]	8,56	8,12	5,35	1,33	2,91
moment obrotowy silnika M [Nm]	42,80	36,33	23,72	1,48	2,67
względne położenie przepustnicy φ_% [%]	14,76	12,94	12,87	2,78	11,09
emisja CO ₂ m_{CO₂} [g/km]	270,98	191,86	410,66	11,31	215,91

Obliczone wartości statystyk opisowych wskazują, że zmienne *prędkość liniowa pojazdu*, *względne obciążenie silnika*, *prędkość obrotowa silnika* charakteryzuje zbliżenie średniej do mediany, co oznacza, że w zbiorze danych nie występują obserwacje odstające, świadczące o braku symetryczności rozkładu. Ponadto, obliczona wartość skośności i kurtozy wskazują na rozkład tych zmiennych zbliżony do rozkładu normalnego. Dla pozostałych zmiennych wartości te wskazują na asymetryczność rozkładu. Dlatego, w kolejnym kroku przystąpiono do weryfikacji zgodności rozkładów badanych zmiennych z rozkładem normalnym. W tym celu wykorzystano test Kołmogorowa-Smirnowa. Dla przyjętego poziomu istotności $\alpha = 0,05$, obliczono wartość statystyki testowej D (19) dla poszczególnych zmiennych oraz odpowiadający jej poziom p . Wartość p dla zmiennych *prędkość liniowa pojazdu*, *względne obciążenie silnika* oraz *prędkość obrotowa silnika* nie daje podstaw do odrzucenia hipotezy zerowej, zakładającej, że badane próby pochodzą z populacji o rozkładzie normalnym. W odniesieniu do zmiennych *moc silnika*, *moment obrotowy silnika* oraz *emisja CO₂* otrzymane wyniki nie pozwoliły na przyjęcie tej hipotezy (tabela 13).

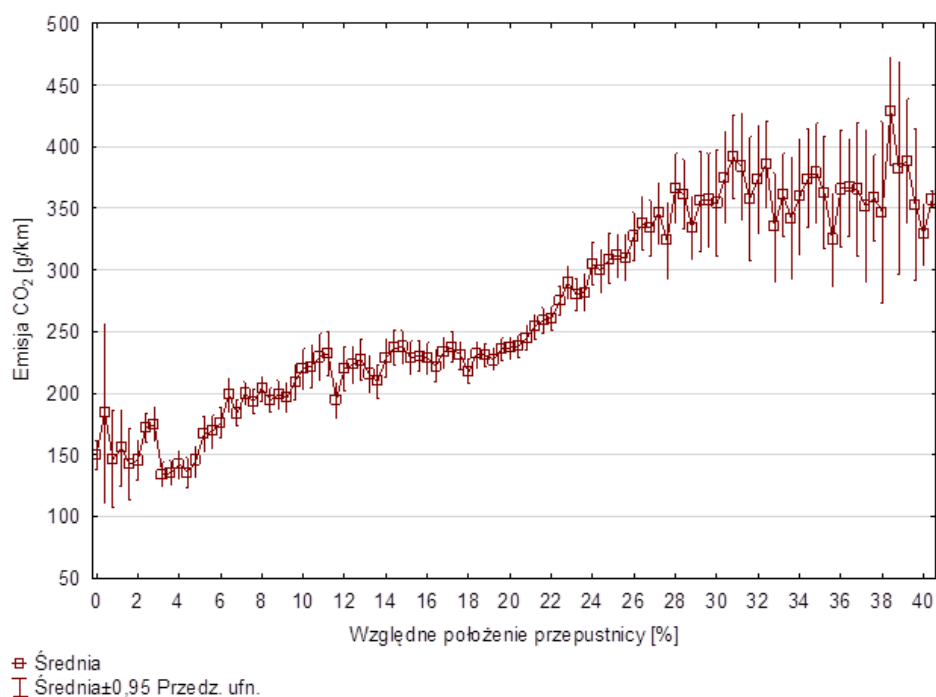
Tabela 13. Wyniki testu Kołmogorowa-Smirnowa

Badana zmienna	Wartość statystyki testowej D	Wartość p
prędkość pojazdu v [km/h]	0,007	0,08
względne obciążenie silnika $L_{\%}$ [%]	0,008	0,10
prędkość obrotowa silnika n [obr/min]	0,004	0,12
moc silnika P [kW]	0,111	0,01
moment obrotowy silnika M [Nm]	0,174	0,00
względne położenie przepustnicy $\varphi_{\%}$ [%]	0,145	0,01
emisja CO ₂ m_{CO_2} [g/km]	0,276	0,01

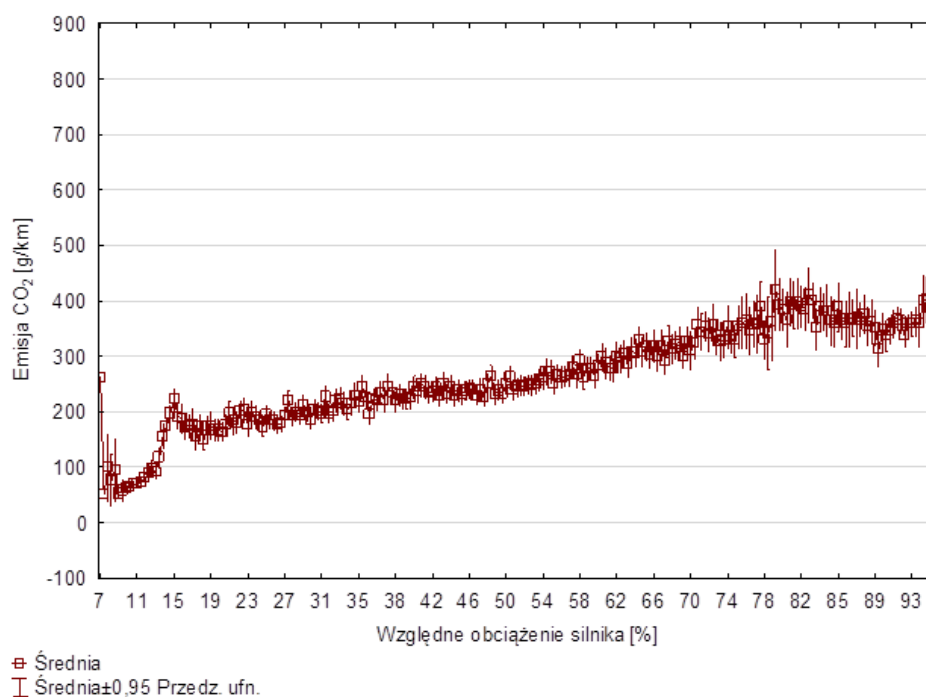
W następnym kroku sprawdzono liniowość przebiegu wytypowanych zmiennych zależnych względem zmiennej emisja CO₂ (rys. 35–41). Rozpoczęto od *prędkości pojazdu*.

Rys. 35. Wykres emisji CO₂ w zależności od *prędkości pojazdu*

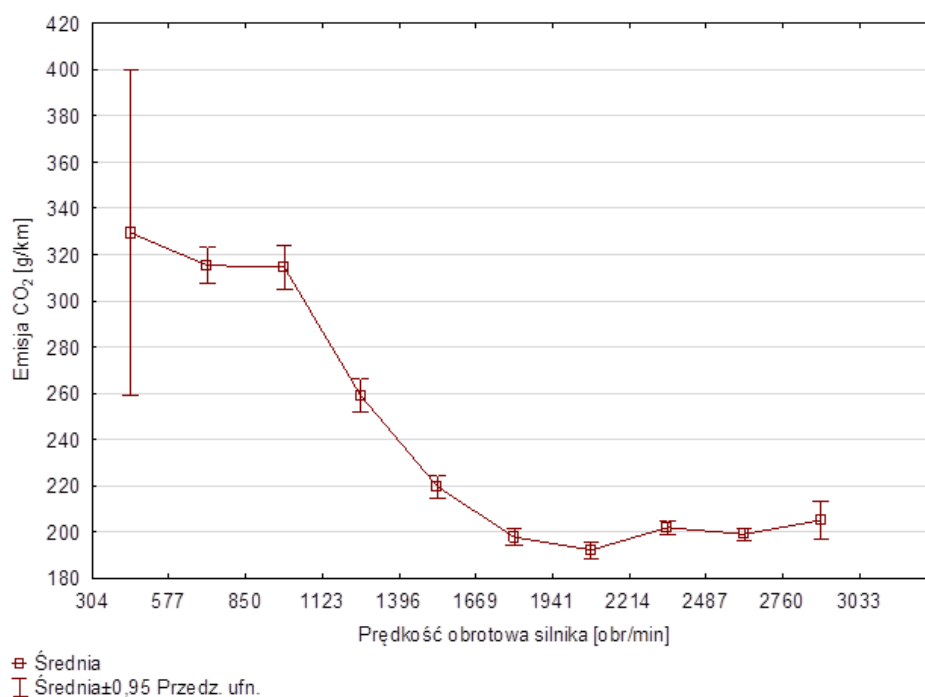
Powyższa analiza wykazała, że poziom emisji CO₂ znacząco zmienia się w zależności od *prędkości pojazdu*. Średnia emisja CO₂ przy niskich prędkościach jest wyższa niż przy prędkościach wyższych. Następnie wykreślono wykresy przebiegu zmiennych względne położenie przepustnicy, względne obciążenie silnika oraz prędkość obrotowa (rys. 36–38).



Rys. 36. Wykres *emisji CO₂* w zależności od *względnego położenia przepustnicy*



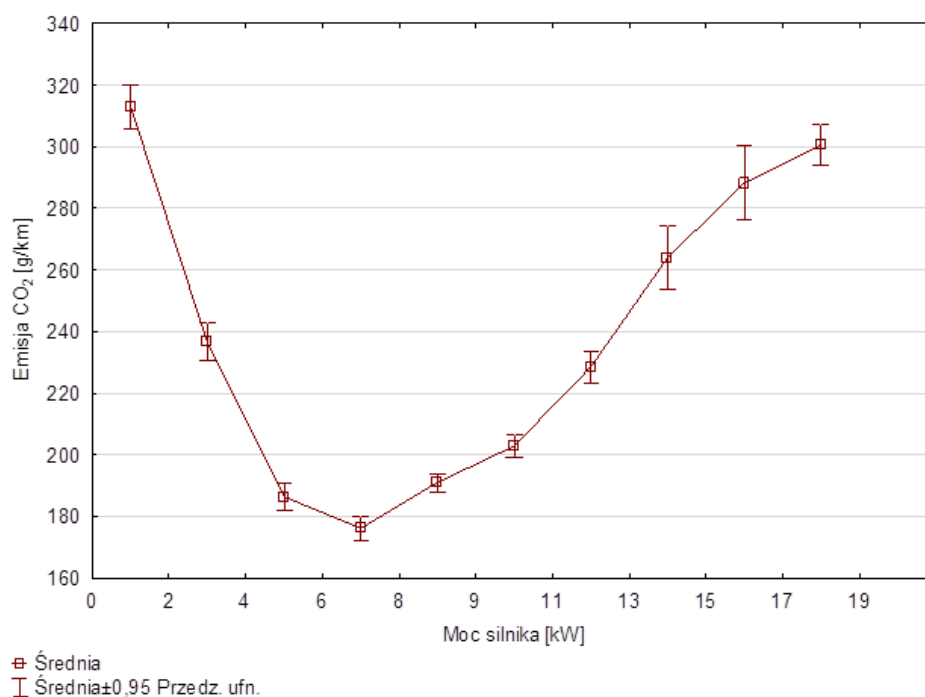
Rys. 37. Wykres *emisji CO₂* w zależności od *względnego obciążenia silnika*



Rys. 38. Wykres *emisji CO₂* w zależności od *prędkości obrotowej*

Zauważyć należy, że w przypadku wszystkich trzech zmiennych widoczna jest zależność względem *emisji CO₂*. Dla zmiennych względne *położenie przepustnicy* oraz względne *obciążenie silnika* wartość *emisji CO₂* rośnie wraz z ich wzrostem natomiast dla zmiennej *prędkość obrotowa*. Potwierdza to zasadność uwzględnienia badanych zmiennych w modelu regresji wielorakiej.

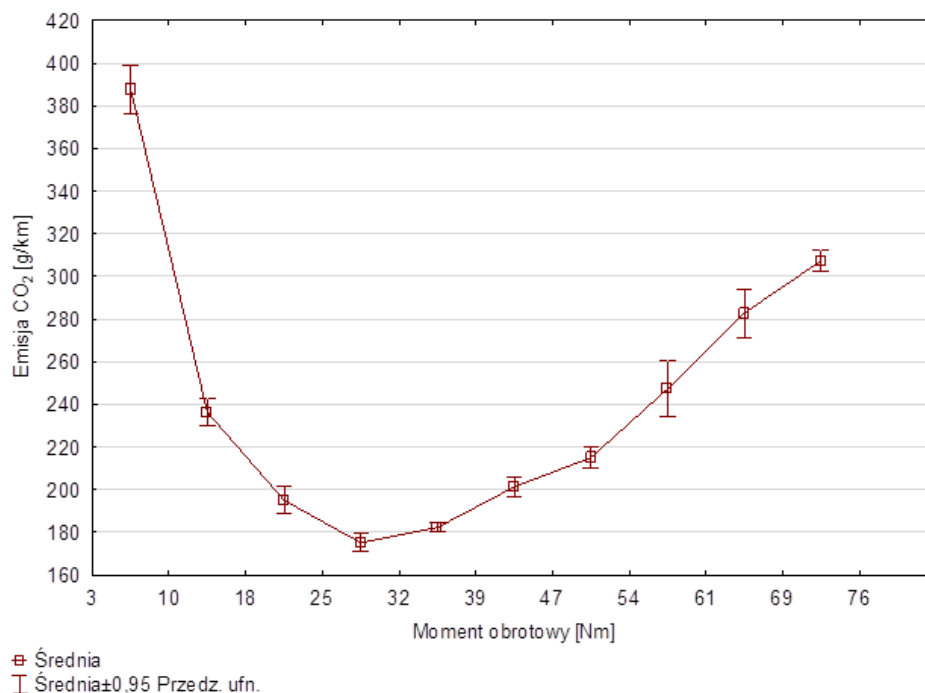
Jako kolejny wykreślono wykres zmiennej *moc silnika* w zależności od *emisji CO₂* (rys. 39–41).



Rys. 39. Wykres *emisji CO₂* w zależności od *mocy silnika*

Zauważyć należy, że *emisja CO₂* jest funkcją niemonotoniczną w zależności od wartości mocy silnika. Do wartości 7,5 kW jest funkcją malejącą, następnie przechodzi w funkcję rosnącą. Zatem charakteryzuje ją nieliniowy przebieg względem zmiennej *emisja CO₂*.

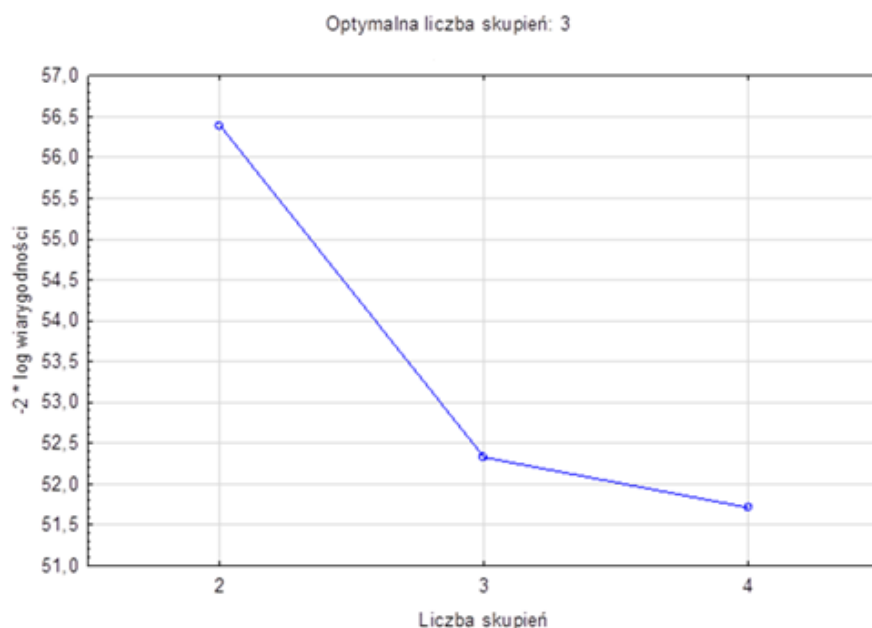
Jako kolejny wykreślono wykres zmiennej *moment obrotowy* w zależności od *emisji CO₂* (rys. 40).



Rys. 40. Wykres *emisji CO₂* w zależności od *momentu obrotowego*

W przypadku momentu obrotowego emisja CO₂ jest również funkcją niemonotoniczną, malejącą do wartości momentu obrotowego wynoszącego ok. 30 Nm, a następnie przechodzącą w funkcję rosnącą.

Ze względu na obserwowaną dużą zmienność, w dalszej analizie dokonano podziału obserwacji na grupy podobne. Wykorzystano w tym celu metodę *k*-średnich. W pierwszym kroku, stosując metodę łokcia, wyznaczono optymalną liczbę skupień (rys. 41), a następnie, wyodrębniono 3 jednorodne grupy obserwacji w badanej populacji.



Rys. 41. Wykres optymalnej liczby skupień z zastosowaniem metody łokcia

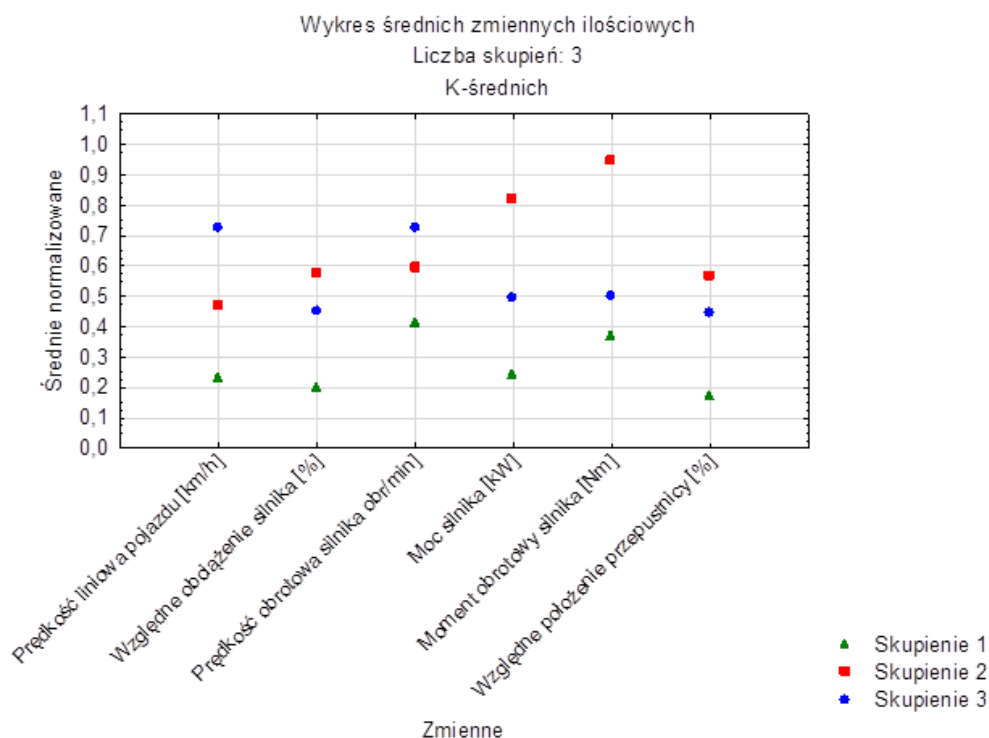
Średnie wartości poszczególnych zmiennych w wyodrębnionych grupach przedstawiono w tabeli 14 oraz na rysunku 42, natomiast ich liczności w tabeli 15.

Tabela 14. Średnie wartości zmiennych w poszczególnych skupieniach

Numer skupienia	Prędkość liniowa pojazdu v [km/h]	Względne obciążenie silnika $L_{\%}$ [%]	Prędkość obrotowa silnika n [obr/min]	Moc silnika P [kW]	Moment obrotowy silnika M [Nm]	Względne położenie przepustnicy $\varphi_{\%}$ [%]	Emisja CO ₂ m_{CO_2} [g/km]
1	16,78	24,65	1411,49	4,41	28,70	6,95	365,60
2	105,68	42,48	2252,47	9,85	41,82	15,93	166,25
3	49,17	66,64	1910,11	14,36	67,70	26,03	213,40

Tabela 15. Liczebność obserwacji w skupieniach

Numer skupienia	Liczba obserwacji	Udział (%)
1	3846	16,86
2	9318	40,85
3	9647	42,29



Rys. 42. Średnie wartości zmiennych w poszczególnych skupieniach

Powyższa analiza przebiegu funkcji emisji w zależności od wytypowanych parametrów wykazała, że dostępną bazę danych należy podzielić na trzy grupy, związane z pracą silnika tj.: tryb ruszania pojazdu (skupienie 1 – *R*), tryb jazdy miejskiej (skupienie 3 – *M*) oraz tryb jazdy pozamiejskiej (skupienie 2 – *T*), dla których średnia wartość prędkości pojazdu wynosi odpowiednio 16 km/h, 49 km/h oraz 105 km/h. Przyjąć należy, że tryb ruszania pojazdu obejmuje prędkości pojazdu z zakresu do ok. 35 km/h, tryb jazdy miejskiej do ok. 85 km/h, zaś tryb jazdy pozamiejski prędkości powyżej 85 km/h.

Aby potwierdzić zasadność powyżej zaproponowanego podziału, przeprowadzono badanie jednorodności w wytypowanych podgrupach. Dobór testu poprzedzono analizą przebiegów rozkładów zmiennej *emisja CO₂* w każdej z grup. Jako pierwsze sprawdzono założenie o zgodności z rozkładem normalnym, przez wykonanie testu Kołomogorova-Smirnova, w ramach którego obliczono wartość statystyki testowej *D* (19). Wyniki przedstawiono w tabeli 16.

Tabela 16. Wyniki testu Kołomogorova-Smirnowa na zgodność zmiennej *emisja CO₂* z rozkładem normalnym

Tryb jazdy	Wartość statystyki testowej	Wartość <i>p</i>
R	0,13	0,01
M	0,04	0,01
T	0,12	0,00

W związku z brakiem spełnienia założeń do przeprowadzenia założeń analizy ANOVA, do porównania zmiennej w trzech grupach wykorzystano test Kruskala-Wallisa. Obliczoną wartość statystyki testowej $T(40)$ oraz odpowiadającą jej wartość p przedstawiono w tabeli 17.

Tabela 17. Wyniki testu Kruskala-Wallisa

Kruskala-Wallisa: T=5631,46 p =0,00		
Tryb jazdy	Suma rang	Średnia ranga
R	71 268 902	18 516,21
M	103 018 512	10 673,28
T	85 894 852	9 226,09

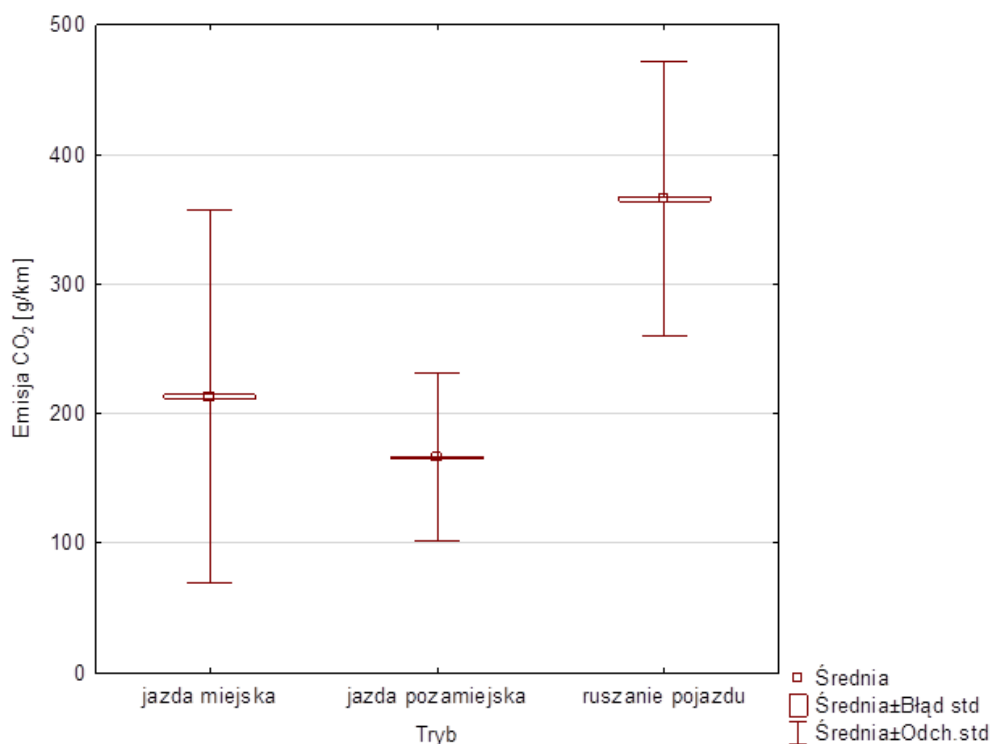
Obliczona wartość p jest mniejsza od przyjętego poziomu istotności $\alpha = 0,05$, w związku z czym należy odrzucić hipotezę zerową i przyjąć twierdzenie o istnieniu różnicy średnich wartości w poszczególnych grupach.

Następnie przeprowadzono test post hoc Tukey'a z poprawką Kramera w celu sprawdzenia istotności różnic średnich wartości *emisji CO₂* w poszczególnych skupieniach. Obliczono wartości statystyk testowych q (42) oraz odpowiadającą im wartość p (tabela 18), które nie pozwoliły na przyjęcie hipotezy zerowej o tym że porównywane średnie grup nie różnią się statystycznie.

Tabela 18. Wyniki testu Tukey'a-Kramera

Test Tukey'a-Kramera HSD (nierówne N) zmienna <i>Emisja CO₂</i>			
Tryb jazdy	R 365,60 g/km	M 213,40 g/km	T 166,25 g/km
R		0,00	0,00
M			0,00
T	0,00	0,00	

Stwierdzone różnice widoczne są na rysunku 43, na którym przedstawiono wykres ramkowy *emisji CO₂* w zależności od trybu jazdy.



Rys. 43. Wykres średnich emisji CO_2 w zależności od trybu jazdy

5.2. Modelowanie emisji CO_2 z zastosowaniem modelu regresji wielorakiej

5.2.1. Model regresji wieloczynnikowej emisji CO_2 dla ruszania pojazdu

W kolejnym etapie badania przystąpiono do budowy modelu emisji CO_2 do środowiska w momencie ruszania pojazdu z zastosowaniem modelu regresji wieloczynnikowej. Jako zmienne niezależne przyjęto *prędkość liniową pojazdu, względne obciążenie silnika, prędkość obrotową silnika, moc silnika, moment obrotowy, względne położenie przepustnicy*.

W pierwszym kroku przeprowadzono analizę danych z zastosowaniem wybranych statystyk opisowych zarejestrowanych w początkowej fazie ruchu pojazdu (tabela 19). Zbadano wartości średniej i mediany oraz odchylenia standardowego (12-14).

Tabela 19. Podstawowe wartości statystyk opisowych zmiennych w skupieniu ruszanie pojazdu

Badana zmienna	Średnia	Mediana	Odchylenie standardowe
prędkość pojazdu [km/h]	16,78	17,00	1,00
względne obciążenie silnika [%]	28,65	25,10	7,06
prędkość obrotowa silnika [obr/min]	1 130,42	1 121,50	304,25
moc silnika [kW]	4,06	2,82	0,07
moment obrotowy [Nm]	32,15	24,69	0,46
względne położenie przepustnicy [%]	7,53	5,49	0,39
emisja CO_2 [g/km]	365,85	360,01	105,82

Średnia wartość prędkości liniowej pojazdu w czasie ruszania pojazdu wynosi 16 km/h przy średniej wartości emisji CO₂ wynoszącej 365 g/km. Obliczone wartości odchylenia standardowego dla zmiennej *emisja CO₂* oraz *prędkość obrotowa silnika* wskazują na wysoką zmienność tych danych.

Następnie przeprowadzono analizę korelacji wszystkich badanych zmiennych. W tym celu zastosowano współczynnik korelacji Pearsona zgodnie ze wzorem 54 (tabela 20).

Tabela 20. Współczynniki korelacji zmiennych objaśniających i objaśnianej w trybie ruszania pojazdu

Badana zmienna	prędkość pojazdu [km/h]	względne obciążenie silnika [%]	prędkość obrotowa silnika [obr/min]	moc silnika [kW]	moment obrotowy [Nm]	względne położenie przepustnicy [%]	emisja CO ₂ [g/km]
prędkość pojazdu [km/h]	1,00	0,10	0,84	0,50	0,34	0,27	-0,37
względne obciążenie silnika [%]		1,00	0,15	0,84	0,47	0,43	0,61
prędkość obrotowa silnika [obr/min]			1,00	0,72	0,33	0,29	-0,16
moc silnika [kW]				1,00	0,94	0,45	0,03
moment obrotowy [Nm]					1,00	0,45	0,07
względne położenie przepustnicy [%]						1,00	0,51

Przeprowadzona analiza wykazała, że wartość współczynnika korelacji pomiędzy zmienną *prędkość pojazdu* a *prędkość obrotowa silnika* wynosi 0,84. Równie wysoką wartość obliczono dla zmiennych *względne obciążenie silnika* i *moc silnika* (0,84). Wysokie wyniki współczynnika korelacji uzyskano dla zmiennych *moc silnika* oraz *moment obrotowy* (0,94) i *prędkość obrotowa silnika* (0,72). Dlatego, w dalszej analizie nie uwzględniono zmiennych *prędkość obrotowa silnika*, *moment obrotowy* oraz *moc silnika*. Przed przystąpieniem do budowy modelu zbiór danych podzielono na dwie próby, stanowiące odpowiednio 90% i 10% całego zbioru, które następnie wykorzystano odpowiednio jako próbę uczącą oraz testową. Taki podział umożliwi ocenę czy model osiąga satysfakcjonujące wyniki prognoz także dla

danych, które nie zostały wykorzystane do jego budowy i pozwoli zweryfikować jego zdolność do generalizacji oraz potwierdzi użyteczność w zastosowaniach praktycznych.

W pierwszym kroku przeprowadzono estymację parametrów modelu, które następnie sprawdzono pod kątem istotności statystycznej. Przeprowadzono test t-Studenta, w ramach którego obliczono wartość statystyki testowej t oraz odpowiadającą jej wartość p (49). Dla wszystkich zmiennych wartość $p=0,00$, co oznacza, że nie ma podstaw do przyjęcia hipotezy zerowej. Obliczone wartości parametrów są istotnie różne od zera. Oszacowane wartości parametrów wraz z wartością statystyki testowej oraz otrzymanym poziomem prawdopodobieństwa p przedstawiono w tabeli 21.

Tabela 21. Oszacowane wartości parametrów modelu regresji wieloczynnikowej w trybie ruszania pojazdu

N=3475	Podsumowanie regresji zmiennej zależnej: <i>Emisja CO₂</i> R ² = 0,70 F(3,3471)=1164,4, p<0,00 Błąd standardowy estymacji: 74,74 g/km		
	b	$t(3471)$	wartość p
Wyraz wolny	358,72	71,55	0,00
Prędkość pojazdu	-2,59	-57,94	0,00
Względne obciążenie silnika	3,06	19,40	0,00
Względne położenie przepustnicy	-0,75	-2,48	0,01

W następnej kolejności dokonano oceny modelu. W tym celu w pierwszej kolejności obliczono wartość współczynnika determinacji R^2 , który wynosi 70%. Współczynnik ten wskazuje, że model w 70% wyjaśnia zmienność zmiennej *emisja CO₂*, co należy uznać za dobry wynik.

Sprawdzono również łączną istotność parametrów strukturalnych. Obliczona wartość statystyki testowej (46) w uogólnionym teście Walda wynosi $F = 1164,4$, a odpowiadający jej poziom prawdopodobieństwa $p=0,00$, co oznacza, że wszystkie współczynniki regresji w modelu są w sposób istotny różne od zera.

Na tej podstawie otrzymano równanie modelu regresji:

$$y = 358,71 - 2,59 \cdot V + 3,06 \cdot L_{\%} - 0,75 \cdot \varphi_{\%},$$

gdzie:

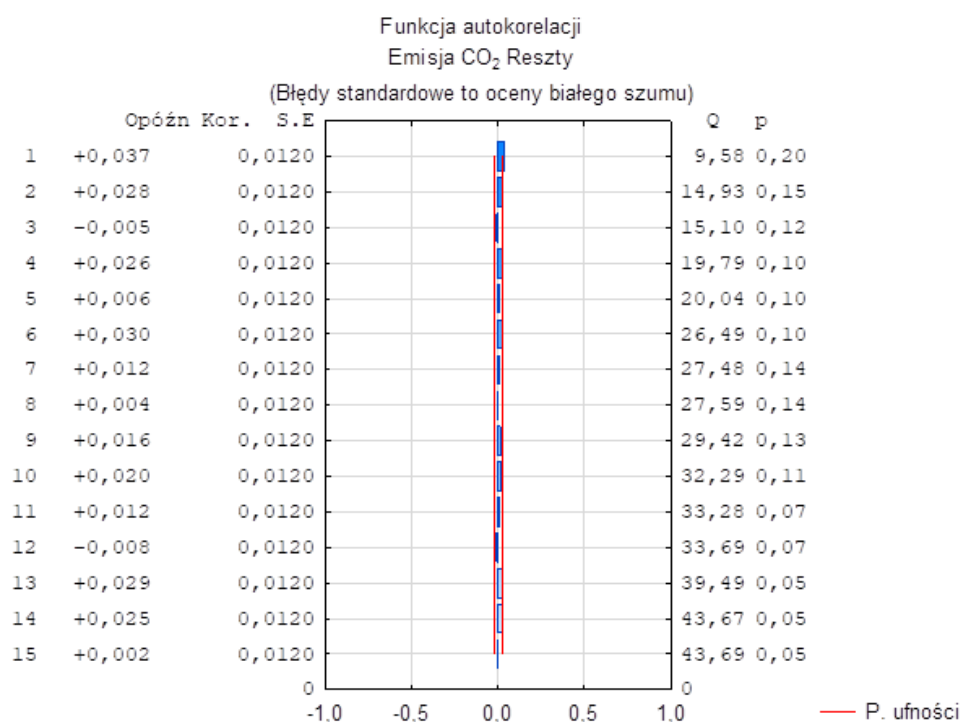
V – prędkość pojazdu,

$L_{\%}$ – względne obciążenie silnika,

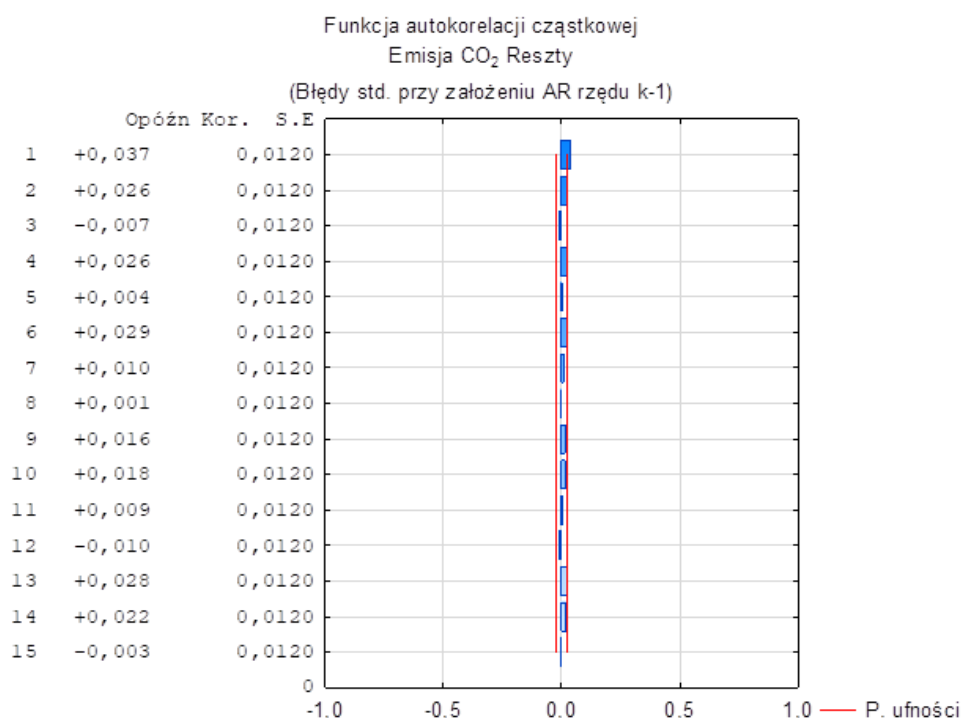
$\varphi_{\%}$ – położenie przepustnicy.

Następnie dokonano badania rozkładu reszt, najpierw oceniając zgodność z rozkładem normalnym. Obliczona wartość statystyki testowej (57) testu Chi-kwadrat wynosi 267,18, a odpowiadający jej poziom prawdopodobieństwa $p=0,09$, wskazuje, że nie ma podstaw do odrzucenia hipotezy zerowej o zgodności rozkładu reszt z rozkładem normalnym.

W celu sprawdzenia założenia o braku autokorelacji reszt wykreślono wykresy funkcji autokorelacji składnika losowego oraz funkcji autokorelacji cząstkowej (rys. 44–45).



Rys. 44. Wykres funkcji autokorelacji składnika resztowego modelu w trybie ruszania pojazdu



Rys. 45. Wykres funkcji autokorelacji cząstkowej składnika resztowego modelu w trybie ruszania pojazdu

Na obydwu wykresach nie pojawiają się statystycznie istotne wartości funkcji autokorelacji, co oznacza, że rozkład reszt modelu jest losowy.

Przeprowadzono również test Durбина-Watsona. Obliczona wartość statystyki testowej (62) wynosi $DW = 1,925$, natomiast wartości progowe wynoszą $dL = 1,91$ oraz $dU = 1,92$. Wartość statystyki testowej jest mniejsza od dwóch oraz większa niż górna wartość progowa, co pozwala wnioskować o braku autokorelacji składnika resztowego.

Następnie zbadano heteroskedastyczność rozkładu reszt, wykorzystując test White'a. Obliczona wartość statystyki testowej (59) wynosi $F = 38,81$ oraz odpowiadający jej poziom prawdopodobieństwa $p=0,08$. Powyższe oznacza, że nie ma podstaw do odrzucenia hipotezy zerowej i pozwala przyjąć, że nie występuje zmienność składnika losowego dla poszczególnych zmiennych objaśniających, a zatem spełnione jest założenie o jego homoskedastyczności (wariancja reszt jest jednorodna).

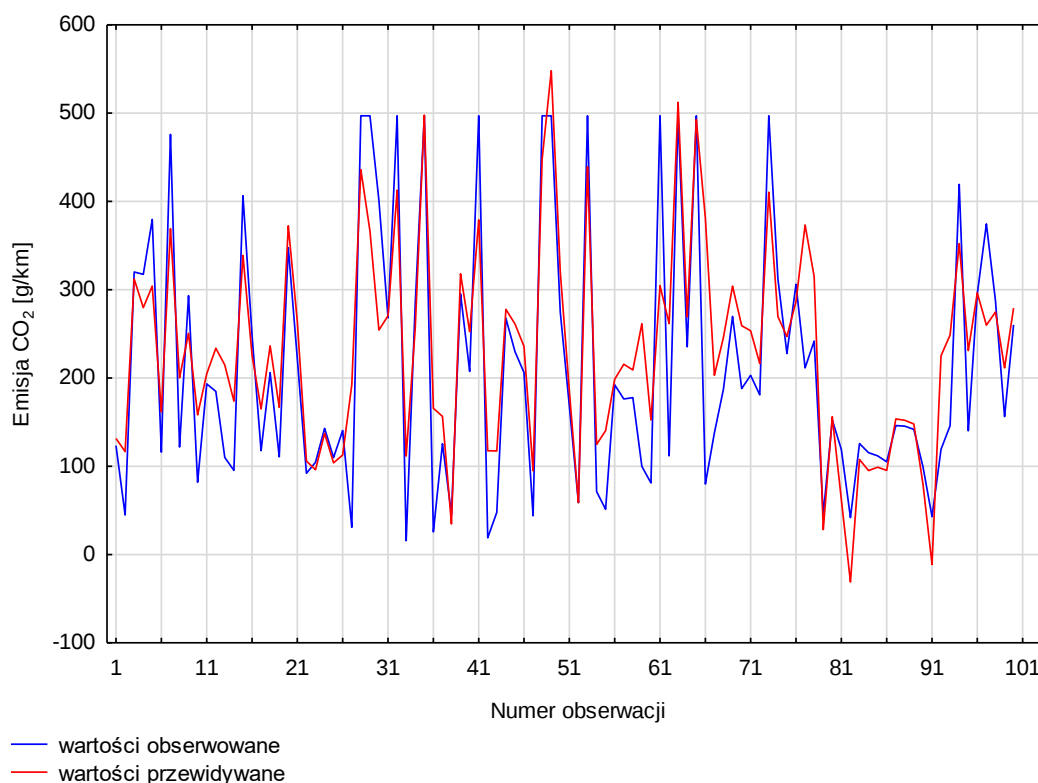
Powyższe analizy potwierdzają poprawność modelu, w związku z czym w kolejnym etapie przeprowadzono ocenę jakości dopasowania. W tym celu wykorzystano zbiór danych testowych, wydzielonych z tej samej próby, co zbiór danych uczących. Umożliwiło to weryfikację czy model cechuje się odpowiednimi i zadowalającymi właściwościami predykcyjnymi. Obliczono kluczowe miary dopasowania, które przedstawiono w tabeli 22.

Tabela 22. Oszacowane wartości błędów prognozy oraz kryterium informacyjnego w trybie ruszania pojazdu

Nazwa miernika	Wartość miernika
Średnia kwadratów reszt (MSE)	4 452,71
Średni błąd bezwzględny	37,79
Względny średni błąd kwadratowy	27,60
Względne odchylenie przeciętne (MRAE)	0,48
Średni absolutny błąd prognozy (MAPE)	0,20
Kryterium AIC	39 373,26

Średni bezwzględny błąd w modelu wynosi 27,60 g/km informując, jaka jest średnia odległość wartości przewidywanych od obserwowanych. Dla średniej wartości *emisji CO₂* w tym skupieniu, wynoszącej 365,85 [g/km], wynik ten należy uznać za dobry. Średni absolutny błąd prognozy kształtuje się na poziomie 20% informując, jaka jest średnia wielkość błędu prognozy, co jest satysfakcjonującym wynikiem. Powyższe wskazuje, że zdolności predykcyjne modelu kształtują się na dobrym poziomie.

Następnie wykreślono wykres wartości obserwowanych i przewidywanych przez model (rys. 46). Wizualna inspekcja pozwala stwierdzić, że model podąża za danymi empirycznymi, odpowiednio reagując na wzrost i spadek wartości przewidywanej.



Rys. 46. Wykres wartości obserwowanych i przewidywanych dla modelu regresji w trybie ruszania pojazdu

5.2.2. Model regresji wieloczynnikowej w trybie jazdy miejskiej

Jako kolejny opracowano model *emisji CO₂* do środowiska w czasie jazdy w trybie miejskim. W pierwszym kroku przeprowadzono analizę danych z zastosowaniem wybranych statystyk opisowych (12–14), obliczając wartość średnią, medianę i odchylenie standardowe (tabela 23).

Tabela 23. Podstawowe statystyki opisowe zmiennych w trybie jazdy miejskiej

Badana zmienna	Średnia	Mediana	Odchylenie standardowe
prędkość pojazdu [km/h]	49,17	49,00	10,64
względne obciążenie silnika [%]	32,37	28,63	18,72
prędkość obrotowa silnika [obr/min]	1 738,33	1 735,75	247,57
moc silnika [kW]	8,23	6,11	5,23
moment obrotowy [Nm]	44,65	32,50	26,50
względne położenie przepustnicy [%]	11,70	8,63	10,14
emisja CO ₂ [g/km]	213,21	175,52	144,24

Średnia wartość prędkości liniowej pojazdu w tym skupieniu wynosi 49 km/h i jest zbliżona do wartości mediany, co oznacza, że nie ma w jej rozkładzie ekstremalnie odstających obserwacji a rozkład jest symetryczny. Średnia emisja CO₂ kształtuje się na poziomie 213 g/km przy medianie 175 g/km oraz odchyleniu standardowym na poziomie 144,24 g/km. Oznacza to, że zmienna ta cechuje się dużą zmiennością.

Następnie przeprowadzono analizę wzajemnego powiązania zmiennych objaśniających i zmiennej objaśnianej. W tym celu zastosowano współczynnik korelacji Pearsona (54). Wyniki przedstawiono w tabeli 24.

Tabela 24. Współczynniki korelacji zmiennych objaśniających i objaśnianych

Badana zmienna	prędkość liniowa pojazdu [km/h]	względne obciążenie silnika [%]	prędkość obrotowa silnika [obr/min]	moc silnika [kW]	moment obrotowy [Nm]	względne położenie przepustnicy [%]	emisja CO ₂ [g/km]
prędkość liniowa pojazdu [km/h]	1,00	0,06	0,25	0,12	0,07	0,12	-0,20
względne obciążenie silnika [%]		1,00	-0,01	0,83	0,58	0,24	0,82
prędkość obrotowa silnika [obr/min]			1,00	0,75	0,07	0,05	0,08

Badana zmienna	prędkość liniowa pojazdu [km/h]	względne obciążenie silnika [%]	prędkość obrotowa silnika [obr/min]	moc silnika [kW]	moment obrotowy [Nm]	względne położenie przepustnicy [%]	emisja CO ₂ [g/km]
moc silnika [kW]				1,00	0,97	0,49	0,47
moment obrotowy [Nm]					1,00	0,52	0,49
względne położenie przepustnicy [%]						1,00	0,69

Obliczono, że współczynnik korelacji przyjmuje wysoką wartość dla zmiennych *moc silnika* oraz *moment obrotowy* (0,97), *względne obciążenie silnika* (0,83) oraz *prędkość obrotowa silnika* (0,75). Dlatego zdecydowano o wyeliminowaniu z dalszych badań zmiennych: *moment obrotowy*, *prędkość obrotowa silnika* oraz *moc silnika*. Zauważyć należy również, że wszystkie wytypowane zmienne cechuje związek liniowy ze zmienną objaśnianą.

Następnie przystąpiono do budowy modelu. Ponownie bazowano na zbiorze danych uczących.

W pierwszym kroku przeprowadzono estymację parametrów modelu, które następnie sprawdzono pod kątem istotności statystycznej. Wykorzystano test t-Studenta, w ramach którego oszacowano wartość statystyki testowej t (49) oraz odpowiadającą jej wartość p . Oszacowane wartości parametrów wraz z wartością statystyki testowej oraz otrzymanym poziomem prawdopodobieństwa p przedstawiono w tabeli 25. Dla wszystkich zmiennych wartość $p=0,00$, co oznacza, że nie ma podstaw do przyjęcia hipotezy zerowej, a zatem oszacowane parametry są istotnie różne od zera.

Tabela 25. Oszacowane parametry modelu regresji w trybie jazdy miejskiej

N=8712	Podsumowanie regresji zmiennej zależnej: <i>Emisja CO₂</i> R ² =0,63; F(3,8708)=4894,9 p<0,00 Błąd standardowy estymacji: 88,02		
	b	$t(8708)$	wartość p
Wyraz wolny	182,84	62,74	0,00
Prędkość liniowa pojazdu	-4,11	-78,18	0,00
Względne obciążenie silnika	4,23	32,40	0,00
Względne położenie przepustnicy	9,02	25,69	0,00

Na tej podstawie przystąpiono do formalnej weryfikacji modelu. Wartość współczynnika determinacji modelu R^2 wynosi 63%, wskazując, że 63% zależności pomiędzy zmiennymi jest wyjaśniane przez model, co pozwala uznać go za zadowalający.

Sprawdzono również łączną istotność parametrów strukturalnych. Obliczona wartość statystyki testowej w uogólnionym teście Walda (46) wynosi $F = 4\,894,9$, a odpowiadający jej poziom prawdopodobieństwa $p=0,00$, co oznacza, że należy odrzucić hipotezę zerową względem alternatywnej, a zatem wszystkie oszacowane parametry w modelu są w sposób istotny różne od zera.

Na tej podstawie otrzymano równanie modelu regresji:

$$y = 182,84 - 4,11 \cdot V + 4,23 \cdot L_{\%} + 9,02 \cdot \varphi_{\%},$$

gdzie:

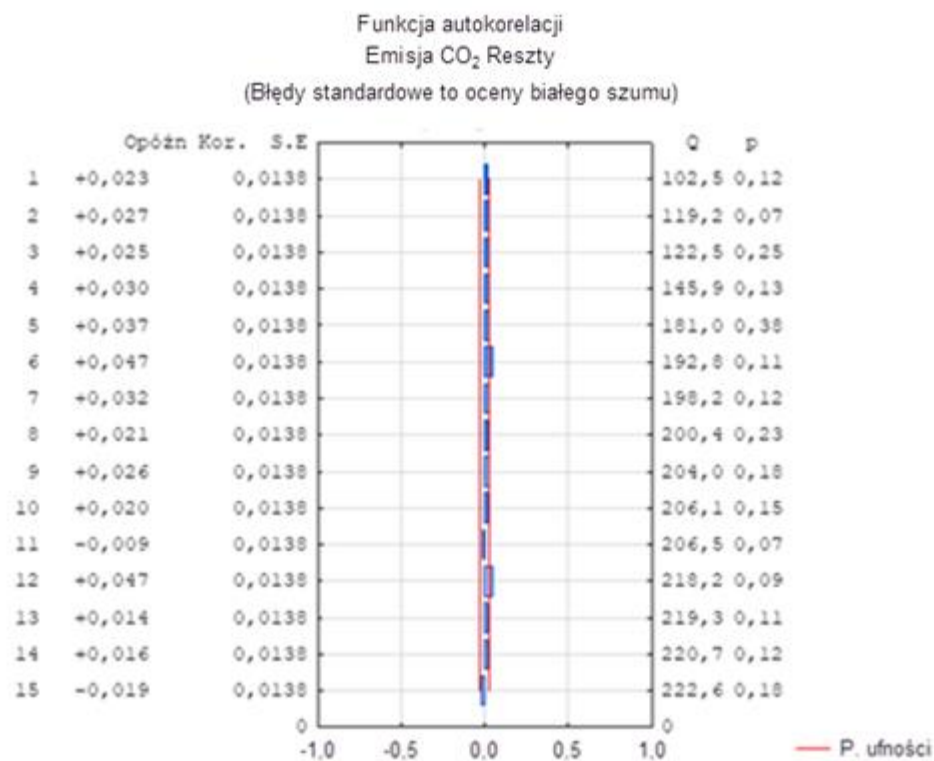
V – prędkość liniowa pojazdu,

$L_{\%}$ – względne obciążenie silnika,

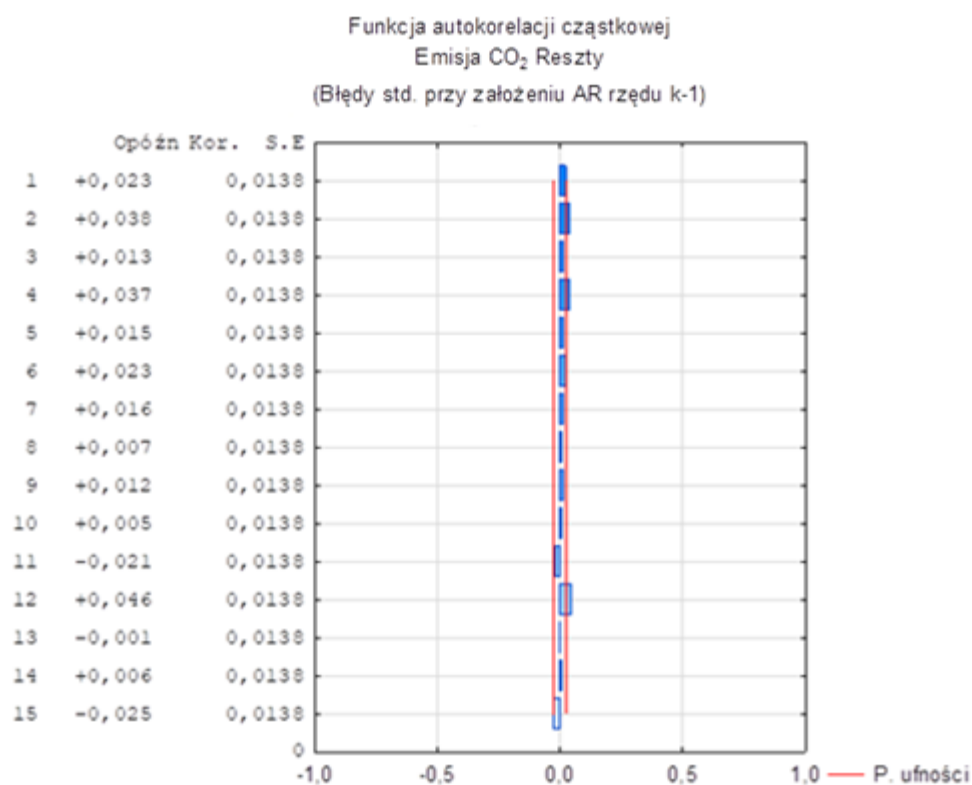
$\varphi_{\%}$ – położenie przepustnicy.

Następnie dokonano badania rozkładu reszt, rozpoczynając od sprawdzenia ich zgodności z rozkładem normalnym. Obliczona wartość statystyki testowej (57) w teście chi-kwadrat wynosi 1761,49, a odpowiadający jej poziom prawdopodobieństwa $p=0,07$, co wskazuje, że nie ma podstaw do odrzucenia hipotezy zerowej, a zatem rozkład reszt jest zgodny z rozkładem normalnym.

Kolejno sprawdzono autokorelację reszt modelu z zastosowaniem testu Durбина-Watsona. Rozpoczęto od wykreślenia wykresów funkcji autokorelacji oraz funkcji autokorelacji cząstkowej składnika losowego (rys. 47–48).



Rys. 47. Wykres funkcji autokorelacji składnika resztowego modelu w trybie jazdy miejskiej



Rys. 48. Wykres funkcji autokorelacji cząstkowej składnika resztowego modelu w trybie jazdy miejskiej

Na obydwu wykresach nie pojawiają się statystycznie istotne wartości funkcji autokorelacji, co oznacza, że rozkład reszt jest losowy.

Następnie obliczono wartość statystyki testowej DW (62) i wyznaczono wartości krytyczne dL oraz dU , przyjmując $k = 6$ (liczba parametrów modelu) oraz $T = 8712$ (liczba obserwacji). Obliczona wartość statystyki testowej $DW = 1,94$, natomiast wartości progowe wynoszą $dL = 1,92$ oraz $dU = 1,93$. Wartość statystyki testowej jest mniejsza od dwóch oraz większa niż górna wartość progowa, co pozwala wnioskować o braku autokorelacji składnika resztowego (nie ma podstaw do odrzucenia hipotezy zerowej).

W kolejnym kroku dokonano sprawdzenia założenia o heteroskedastyczności reszt. Obliczona wartość statystyki testowej w teście White'a wyniosła $F = 111,28$, a odpowiadający jej poziom prawdopodobieństwa $p=0,06$, co oznacza, że nie ma podstaw do odrzucenia hipotezy zerowej i pozwala wnioskować, że składnik losowy jest homoskedastyczny, tj. stała jest wartość jego wariancji dla poszczególnych wartości zmiennych objaśniających.

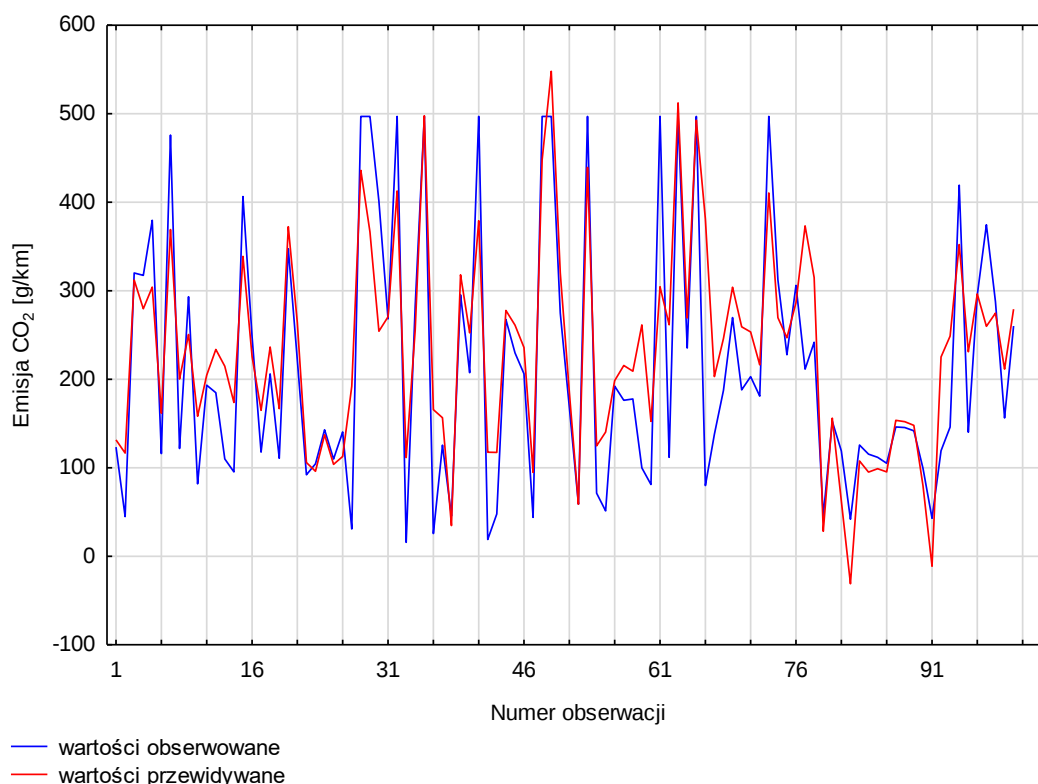
Powyższa analiza pozwoliła na pozytywne zweryfikowanie założeń modelu regresji i uznanie go za poprawny. Dlatego wykorzystując zbiór danych testowych przystąpiono do analizy jakości dopasowania, a następnie oszacowano podstawowe mierniki błędów modelu (tabela 26).

Tabela 26. Oszacowane wartości błędów prognozy oraz kryterium informacyjnego w trybie jazdy miejskiej

Nazwa miernika	Wartość miernika
Średnia kwadratów reszt (MSE)	7 513,39
Średni błąd bezwzględny	60,57
Względny średni błąd kwadratowy	3,45
Względne odchylenie przeciętne (MRAE)	0,52
Średni absolutny błąd prognozy (MAPE)	0,38
Kryterium AIC	10 261,56

Średni bezwzględny błąd modelu wynosi 60,57 g/km informując, ile średnio różni się wartość przewidywana od obserwowanej. Dla średniej wartości emisji CO_2 w tym skupieniu, wynoszącej 213,21 g/km, Oszacowane wartości błędów prognozy oraz kryterium informacyjnego w trybie ruszania pojazdu wynik ten należy uznać za dostateczny. Średni absolutny błąd prognozy kształtuje się na poziomie 38% informując, jaka jest średnia wielkość błędu prognozy, co jest satysfakcjonującym poziomem. Powyższe wskazuje, że zdolności predykcyjne modelu kształtują się na dostatecznym poziomie.

Następnie wykreślono wykres wartości obserwowanych i przewidywanych przez model (rys. 49). Model podąża za danymi, dobrze oddając charakter zmienności emisji CO_2 .



Rys. 49. Wykres wartości obserwowanej i przewidywanej dla modelu w trybie jazdy miejskiej

5.2.3. Model regresji wieloczynnikowej w trybie jazdy pozamiejskiej

Jako kolejny opracowano model *emisji CO₂* do środowiska w czasie jazdy w trybie pozamiejskim z zastosowaniem modelu regresji wieloczynnikowej. Analizę rozpoczęto od obliczenia wybranych statystyk opisowych (12–14), w tym wartości mediany, średniej i odchylenia standardowego (tabela 27).

Tabela 27. Podstawowe statystyki opisowe modelu w trybie jazdy pozamiejskiej

Badana zmienna	Średnia	Mediana	Odchylenie standardowe
prędkość pojazdu [km/h]	105,68	108,00	17,92
względne obciążenie silnika [%]	48,55	47,84	20,24
prędkość obrotowa silnika [obr/min]	2 263,58	2 290,00	311,19
moc silnika [kW]	11,02	9,56	4,60
moment obrotowy [Nm]	46,82	37,60	19,85
względne położenie przepustnicy [%]	20,36	18,43	14,21
emisja CO ₂ [g/km]	166,25	172,53	64,60

Średnia wartość prędkości liniowej pojazdu w trybie miejskim wynosi 105 km/h, przy średniej *emisji CO₂* na poziomie 166 g/km. Wartość średniej *emisji CO₂* jest mniejsza od mediany, co oznacza, że rozkład zmiennej jest lewoskośny. Zauważyć należy wysoką wartość odchylenia standardowego, co świadczy o wysokiej zmienności danych.

Następnie obliczono współczynnik korelacji Pearsona względem zmiennych objaśniających i zmiennej objaśnianej (tabela 28) zgodnie ze wzorem 54.

Tabela 28. Współczynniki korelacji zmiennych objaśniających i objaśnianej w trybie jazdy pozamiejskiej

	prędkość pojazdu [km/h]	względne obciążenie silnika [%]	prędkość obrotowa silnika [obr/min]	moc silnika [kW]	moment obrotowy [Nm]	względne położenie przepustnicy [%]	emisja CO ₂ [g/km]
prędkość pojazdu [km/h]	1,00	0,22	0,89	0,13	-0,15	0,05	0,14
względne obciążenie silnika [%]		1,00	0,20	0,82	0,94	0,38	0,91
prędkość obrotowa silnika [obr/min]			1,00	0,77	-0,14	0,05	0,18
moc silnika [kW]				1,00	0,94	0,48	0,38
moment obrotowy [Nm]					1,00	0,49	0,33
względne położenie przepustnicy [%]						1,00	0,69

Przeprowadzona analiza wykazała, że wartość współczynnika korelacji pomiędzy zmiennymi *prędkość pojazdu* a *prędkość obrotowa silnika* wynosi 0,89. Równie wysoką wartość obliczono dla zmiennych *względne obciążenie silnika* oraz *moment obrotowy* (0,94) i *moc silnika* (0,82). Dlatego z dalszej analizy wyeliminowano zmienne *prędkość obrotowa silnika*, *moment obrotowy* oraz *moc silnika*.

Następnie przystąpiono do estymacji parametrów modelu, zgodnie z przyjętym algorytmem, które sprawdzono pod kątem istotności statystycznej. Wykorzystano test t-Studenta, w ramach którego oszacowano wartość statystyki testowej t (49) oraz odpowiadającą jej wartość p . Oszacowane wartości parametrów wraz z wartością statystyki testowej oraz otrzymanym poziomem prawdopodobieństwa p przedstawiono w tabeli 29. Dla wszystkich zmiennych wartość $p=0,00$, co oznacza, że nie ma podstaw do przyjęcia hipotezy zerowej, a zatem oszacowane parametry są istotnie różne od zera.

Tabela 29. Estymowane parametry modelu regresji wieloczynnikowej w trybie jazdy pozamiejskiej

N=8380	Podsumowanie regresji zmiennej zależnej: <i>Emisja CO₂</i> R ² = 0,78; F(3,8376)=9910,2, p<0,00 Błąd standardowy estymacji: 30,19		
	b	t(8376)	wartość p
Wyraz wolny	56,51	34,70	0,00
Prędkość pojazdu	-0,56	-33,54	0,00
Względne obciążenie silnika	3,21	63,93	0,00
Względne położenie przepustnicy	1,92	14,92	0,00

Następnie przystąpiono do formalnej weryfikacji modelu. Wartość współczynnika determinacji modelu R² wynosi 0,78, zatem 78% zależności pomiędzy zmiennymi zostało wyjaśnione przez model.

Sprawdzono również łączną istotność parametrów strukturalnych z wykorzystaniem testu Fishera-Snedecora. Obliczona wartość statystyki testowej (46) wynosi $F = 9\,910,2$ a odpowiadający jej poziom prawdopodobieństwa $p = 0,00$, co oznacza, że należy odrzucić hipotezę zerową względem alternatywnej, a zatem wszystkie z estymowanych parametrów w sposób istotny różnią się od zera.

Na tej podstawie otrzymano równanie modelu regresji:

$$y = 56,51 - 0,56 \cdot V + 3,21 \cdot L_{\%} + 1,92 \cdot \varphi_{\%},$$

gdzie:

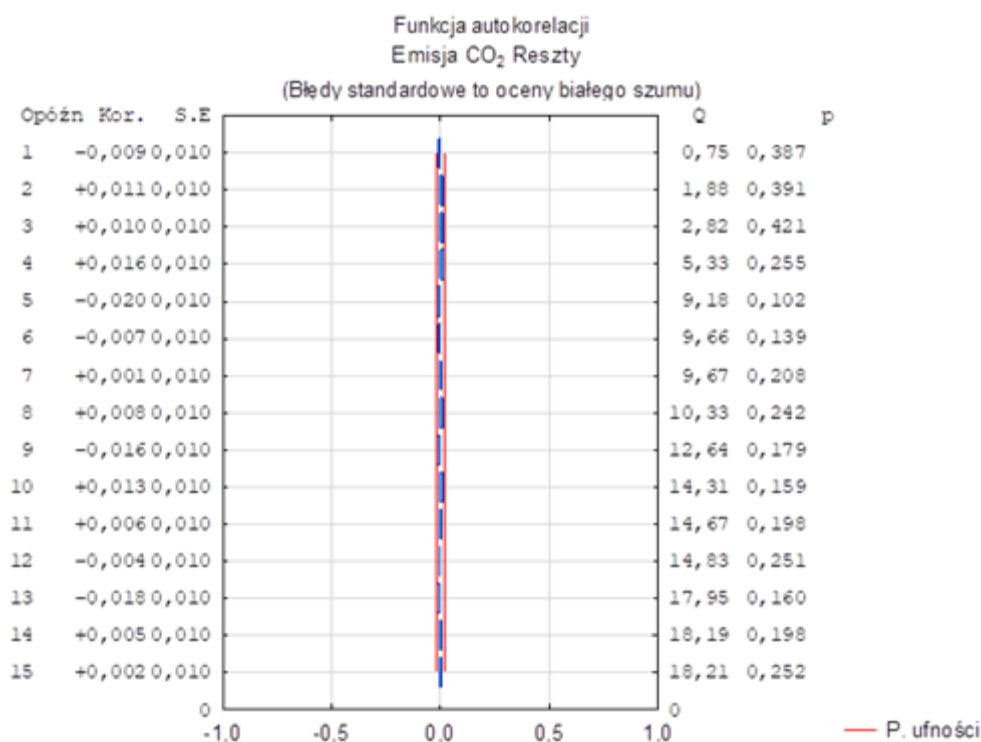
V – prędkość pojazdu,

$L_{\%}$ – względne obciążenie silnika,

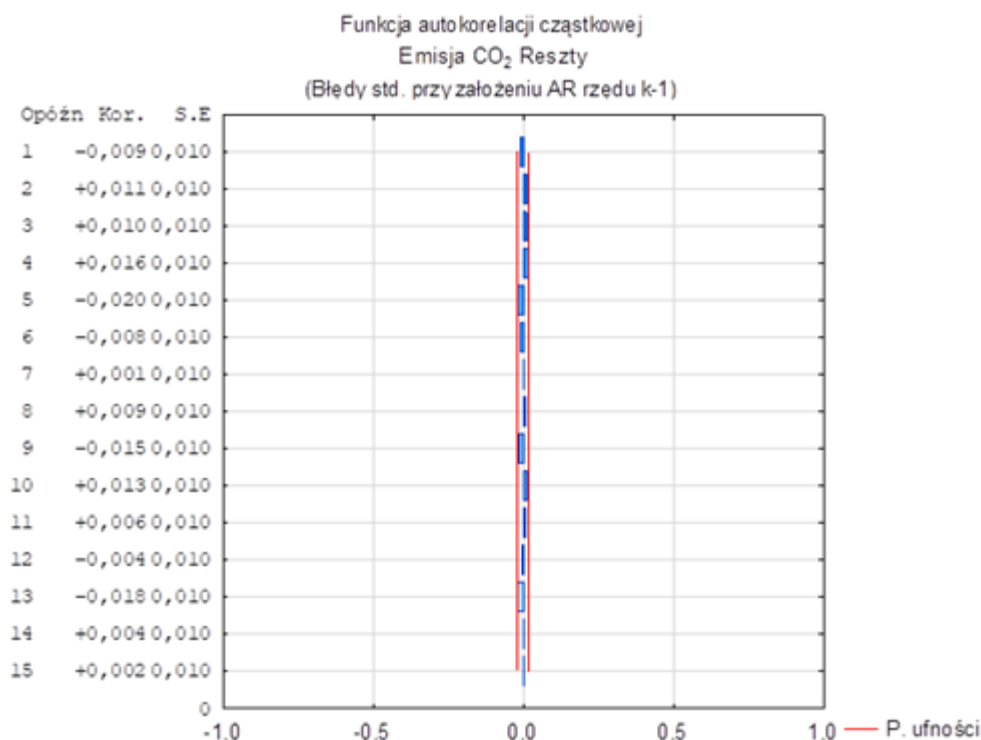
$\varphi_{\%}$ – położenie przepustnicy.

Następnie dokonano analizy rozkładu reszt, rozpoczynając od oceny zgodności ich rozkładu z rozkładem normalnym. Obliczona wartość statystyki testowej (57) w teście chi-kwadrat wynosi 7,28, a odpowiadający jej poziom prawdopodobieństwa $p=0,20$ wskazuje, że nie ma podstaw do odrzucenia hipotezy zerowej, a zatem rozkład reszt jest zgodny z rozkładem normalnym.

W kolejnym kroku sprawdzono występowanie autokorelacji reszt modelu, wykreślając wykresy funkcji autokorelacji składnika losowego oraz funkcji autokorelacji cząstkowej (rys. 50–51).



Rys. 50. Wykres funkcji autokorelacji składnika resztowego modelu w trybie jazdy pozamiejskiej



Rys. 51. Wykres funkcji autokorelacji cząstkowej składnika resztowego modelu w trybie jazdy pozamiejskiej

Na obydwu wykresach nie pojawiają się statystycznie istotne wartości funkcji autokorelacji co oznacza, że rozkład reszt modelu jest losowy. Przeprowadzono również test Durбина-Watsona. Obliczona wartość statystyki testowej (63) $DW = 1,81$, natomiast wartości

progowe wynoszą $dL = 1,79$ oraz $dU = 1,80$. Obliczona wartość statystyki testowej jest większa od dwóch oraz większa niż górna wartość progowa, co pozwala wnioskować o braku autokorelacji składnika resztowego.

W kolejnym kroku przeprowadzono test White'a w celu zbadania występowania heteroskedastyczności reszt. Obliczona wartość statystyki testowej (59) wynosi $F = 39,92$, dla której odpowiadający jej poziom prawdopodobieństwa wynosi $p=0,07$, co oznacza, że nie ma podstaw do odrzucenia hipotezy zerowej i pozwala wnioskować, że składnik losowy jest homoskedastyczny i nie występuje jego zmienność dla poszczególnych zmiennych objaśniających.

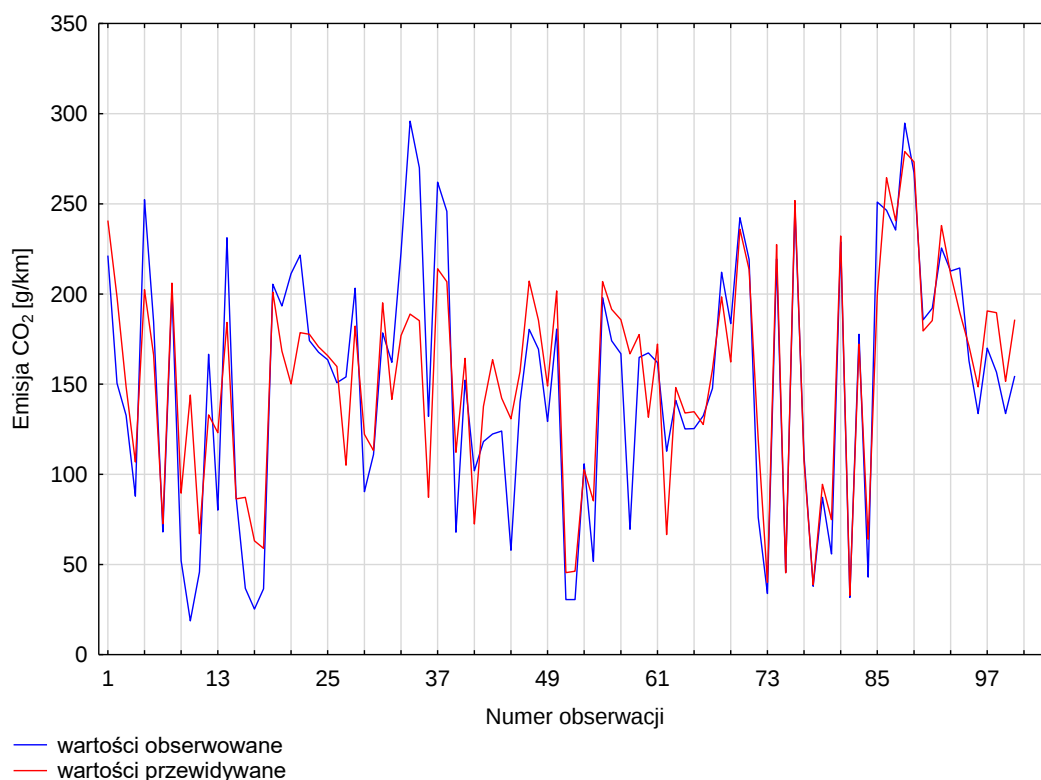
Powyższe pozwoliło wnioskować, że model jest prawidłowy. Po pozytywnej weryfikacji modelu przystąpiono do analizy jakości dopasowania. W pierwszym kroku oszacowano błędy prognozy oraz wartość kryterium AIC (tabela 30).

Tabela 30. Oszacowane wartości błędów prognozy oraz kryterium informacyjnego w trybie jazdy pozamiejskiej

Nazwa miernika	Wartość miernika
Średnia kwadratów reszt (MSE)	1 145,47
Średni błąd bezwzględny	19,62
Względny średni błąd kwadratowy	0,27
Względne odchylenie przeciętne (MRAE)	0,18
Średni absolutny błąd prognozy (MAPE)	0,19
Kryterium AIC	79 686,44

Średni bezwzględny błąd w modelu wynosi 19,62 g/km informując, ile średnio różni się wartość przewidywana od obserwowanej. Dla średniej wartości *emisji CO₂* w tym skupieniu, wynoszącej 166,25 g/km, wynik ten należy uznać za bardzo dobry. Średni absolutny błąd prognozy kształtuje się na poziomie 19% informując, jaka jest średnia wielkość błędu prognozy, co jest satysfakcjonującym wynikiem. Powyższe wskazuje, że zdolności predykcyjne modelu kształtują się na dobrym poziomie.

Następnie wykreślono wykres wartości obserwowanych i przewidywanych (rys. 52) przez model. Jego wizualna inspekcja pozwoliła na stwierdzenie, że zbudowany model bardzo dobrze podąża za danymi, odzwierciedlając spadki i wzrosty modelowanej wartości.



Rys. 52. Wykres wartości obserwowanych i przewidywanych dla modelu regresji w trybie jazdy pozamiejskiej

5.2.4. Porównanie modeli regresji wieloczynnikowej w trzech trybach jazdy

W ostatnim kroku dokonano porównania jakości dopasowania poszczególnych modeli oraz przeprowadzono analizę otrzymanych wyników. W tabeli 31 przedstawiono zestawienie oszacowanych wyrazów wolnych oraz współczynników poszczególnych parametrów trzech modeli. Zauważyć należy, że największy średni poziom *emisji CO₂* do środowiska ma miejsce w trybie ruszania pojazdu (365,60 g/km), natomiast najmniejszy w czasie poruszania się pojazdu w trybie pozamiejskim (166,18 g/km).

Tabela 31. Porównanie parametrów modeli regresji wieloczynnikowej w trzech trybach jazdy

Zmienna objaśniająca	Wartość parametru model w trybie ruszania pojazdu	Wartość parametru model w trybie jazdy miejskiej	Wartość parametru model w trybie jazdy pozamiejskiej
Wyraz wolny	358,72	182,84	56,51
Prędkość pojazdu	-2,59	-4,11	-0,56
Względne obciążenie silnika	3,06	4,23	3,21
Względne położenie przepustnicy	-0,75	9,02	1,92

Niezależnie od trybu jazdy (ruszanie pojazdu, tryb miejski, tryb pozamiejski) każdorazowy wzrost *prędkości jazdy* powoduje zmniejszenie *emisji CO₂* do środowiska. W czasie ruszania pojazdu z każdym wzrostem *prędkości pojazdu* o 1 km/h, *emisja CO₂* spada o 2,59 g/km, podczas jazdy w trybie miejskim spada o 4,11 g/km, natomiast w czasie jazdy w trybie pozamiejskim spada o 0,56 g/km.

W odniesieniu do zmiennej *względne obciążenie silnika*, każdorazowy wzrost wartości analizowanej zmiennej o 1% powoduje wzrost *emisji CO₂* o 3,06 g/km, 4,23 g/km oraz 3,21 g/km odpowiednio dla ruszania pojazdu, jazdy w trybie miejskim oraz w czasie jazdy w trybie pozamiejskim.

Natomiast, w przypadku zmiennej *względne położenie przepustnicy* wraz ze wzrostem wartości tego parametru o 1% podczas ruszania pojazdu zauważyć należy spadek poziomu *emisji CO₂* o 0,75 g/km. Dla jazdy pojazdu w trybie miejskim oraz pozamiejskim, wzrost wartości zmiennej o 1% powoduje wzrost *emisji CO₂* odpowiednio o 9,02 g/km oraz 1,92 g/km.

Tabela 32. Porównanie jakości dopasowania poszczególnych modeli regresji wieloczynnikowej

Model	Błąd estymacji	Średni błąd bezwzględny (MSE)	Względny średni błąd kwadratowy	Względne odchylenie przeciętne (MRAE)	Średni absolutny błąd prognozy (MAPE)	Kryterium AIC
Tryb ruszania pojazdu	74,74	37,79	27,60	0,48	0,20	39 373,26
Tryb jazdy miejskiej	88,02	60,57	3,45	0,52	0,38	10 261,56
Tryb jazdy pozamiejskiej	30,19	19,62	0,27	0,18	0,19	79 686,44

Analiza poszczególnych modeli wykazała (tabeli 32), że największym błędem estymacji cechuje się model regresji w trybie jazdy miejskiej (88,02 g/km), natomiast najmniejszym w czasie jazdy w trybie pozamiejskim (30,19 g/km). Wynika to z faktu, że średnie wartości emisji CO₂ do środowiska, są najwyższe w czasie jazdy pojazdu z niską prędkością. Najmniejszym średnim błędem bezwzględnym cechuje się natomiast model dla jazdy w trybie pozamiejskim (19,62 g/km), natomiast największym model jazdy w trybie miejskim (60,57 g/km).

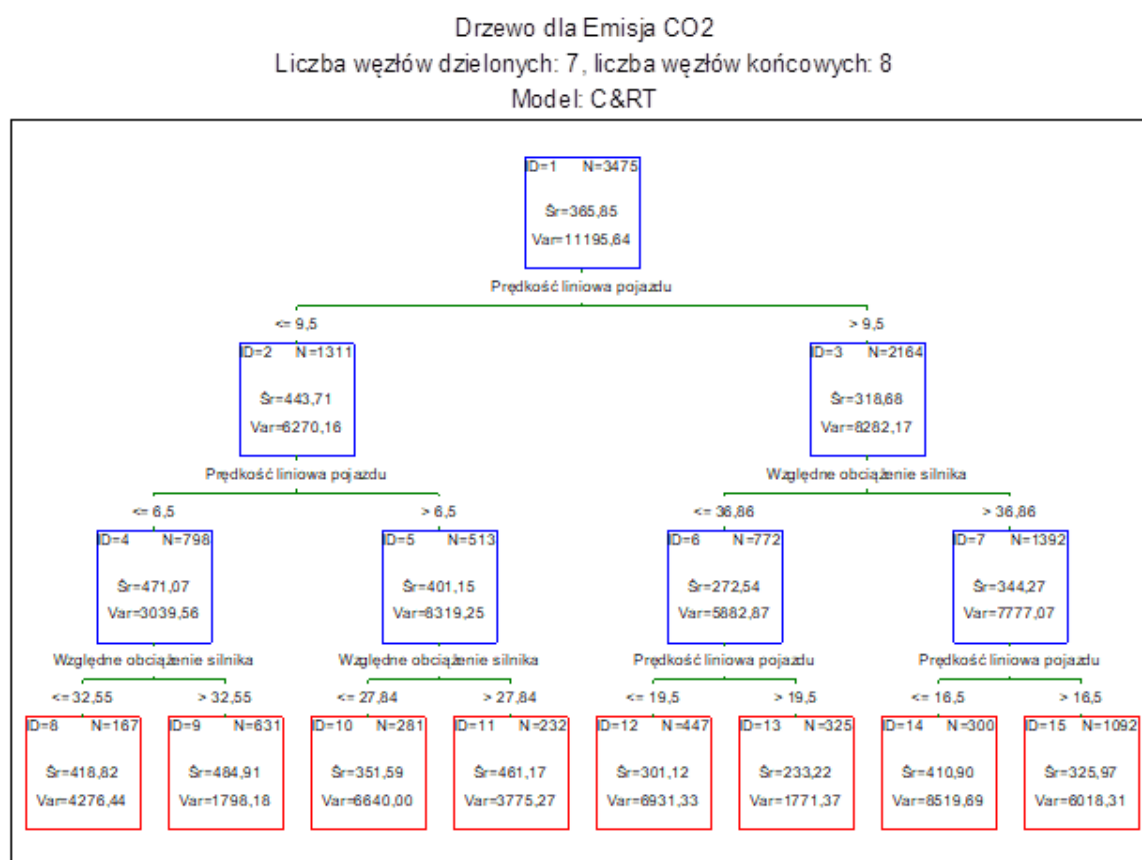
5.3. Modelowanie emisji CO₂ z zastosowaniem modeli drzew decyzyjnych

5.3.1. Modele regresyjnych drzew decyzyjnych dla momentu ruszania pojazdu

W kolejnej części pracy do modelowania i analizy wpływu wytypowanych czynników na poziom *emisji CO₂* zastosowano modele drzew decyzyjnych: z wykorzystaniem algorytmu CART, modele lasu losowego oraz drzew wzmacnianych. Jako zmienne niezależne przyjęto *prędkość pojazdu*, *względne obciążenie silnika*, *względne położenie przepustnicy*. Analizę przeprowadzono w trzech grupach danych obejmujących jazdę zarówno w trybie ruszania, jak i w trybie miejskim i pozamiejskim. Ponownie bazowano na dwóch próbach, testowej oraz uczącej, stanowiących odpowiednio 90%-10% całego zbioru danych. Do budowy modelu wykorzystano próbę uczącą, w ramach której wyodrębniono zbiór walidacyjny, bazując na walidacji 10-krzyżowej. Wszystkie modele zostały skonstruowane zgodnie z założeniami określonymi w algorytmie badania (rys. 33, podrozdział 4.2.1).

Rozpoczęto od budowy modeli dla danych dotyczących trybu ruszania pojazdu. Jako pierwszy zbudowano model drzewa decyzyjnego CART.

Wynik analizy przedstawiono na rysunku 53.



Rys. 53. Drzewo decyzyjne algorytmem CART w trybie ruszania pojazdu

Na podstawie otrzymanych wyników, można wnioskować, że w przypadku jazdy przy *prędkości pojazdu* powyżej 19,5 km/h oraz *względny obciążeniu silnika* poniżej 66,86% poziom emisji CO₂ jest najniższy i przyjmuje wartość średnią 233,22 g/km (liść nr 13). Natomiast, w przypadku jazdy z *prędkością pojazdu* poniżej 6,5 km/h oraz przy *względny obciążeniu silnika* powyżej 42,55% wartość średniej emisji CO₂ jest najwyższy i wynosi 484,91 g/km (liść 9).

W kolejnym kroku zbudowano ranking ważności predyktorów, który umożliwia ocenę wpływu poszczególnych zmiennych objaśniających na badane zjawisko (tabela 33).

Tabela 33. Ranking predyktorów modelu CART w trybie ruszania pojazdu

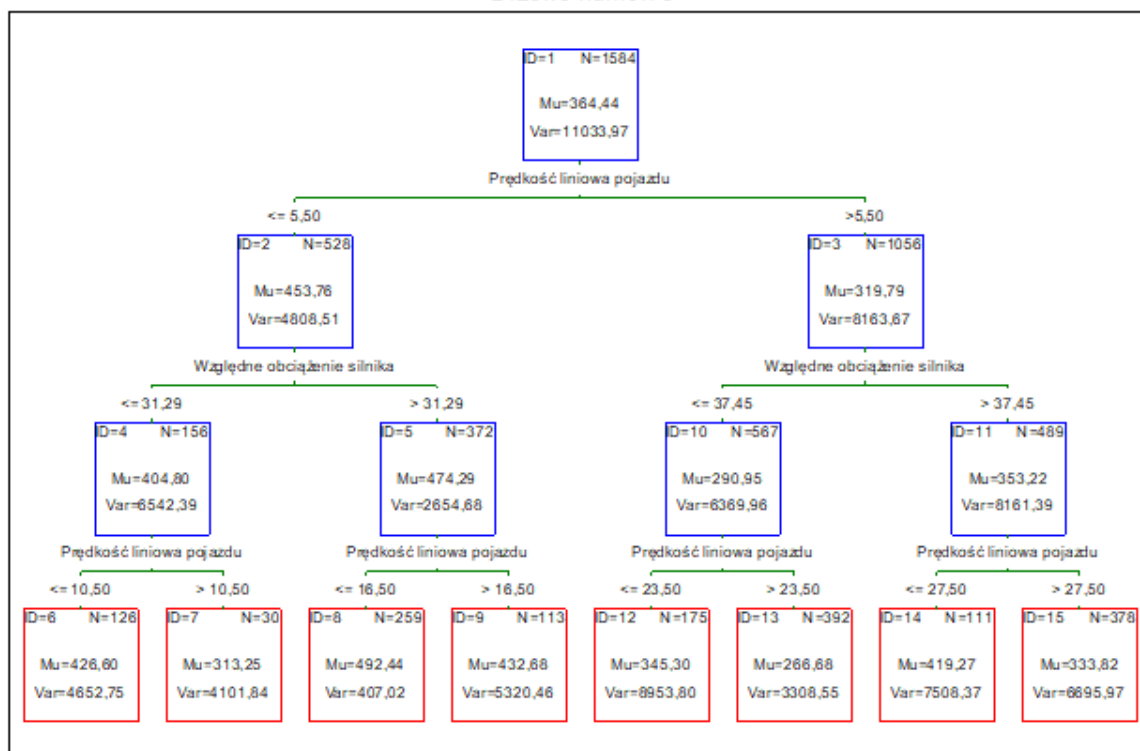
Badana zmienna	Ważność [%]
Prędkość pojazdu [km/h]	100
Względne obciążenie silnika [%]	42
Względne położenie przepustnicy [%]	34

Zauważyć należy, że największy wpływ na badane zjawisko ma *prędkość pojazdu* oraz *względne obciążenie silnika*, natomiast najmniejsze *względne położenie przepustnicy*.

Drugim z opracowywanych modeli był model lasu losowego.

Wynik analizy przedstawiono na rysunku 54. Przedstawia on jedno, przykładowe drzewo wygenerowane w ramach prowadzonego badania.

Drzewo dla Emisja CO2
Liczba węzłów dzielonych: 7, liczba węzłów końcowych: 8
Drzewo numer: 5



Rys. 54. Przykładowe drzewo decyzyjne z zastosowaniem lasu losowego w trybie ruszania pojazdu

Dla opracowanego modelu również zbudowano ranking ważności predyktorów (tabela 34).

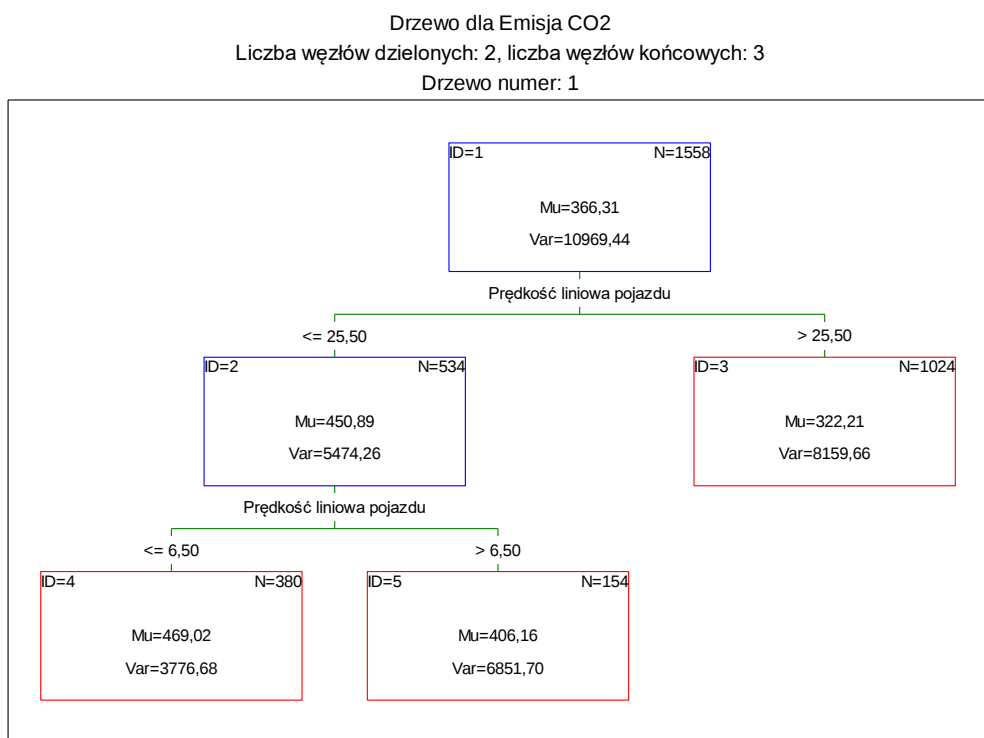
Tabela 34. Ranking predyktorów z zastosowaniem modelu lasu losowego w trybie ruszania pojazdu

Badana zmienna	Ważność [%]
Prędkość pojazdu [km/h]	100
Względne obciążenie silnika [%]	43
Względne położenie przepustnicy [%]	36

Ponownie największy wpływ miała zmienna *prędkość pojazdu* oraz *względne obciążenie silnika*, najmniejszy zaś *względne położenie przepustnicy*.

Jako ostatni opracowano model drzew wzmacnianych (stosując również założenia zdefiniowane w algorytmie badania).

Wynik analizy przedstawiono na rysunku 55.



Rys. 55. Przykładowe drzewo decyzyjne z zastosowaniem drzew wzmacnianych w trybie ruszania pojazdu

Na tej podstawie zbudowano ranking ważności predyktorów dla tego modelu (tabela 35).

Tabela 35. Ranking predyktorów w modelu z zastosowaniem drzew wzmacnianych w trybie ruszania pojazdu

Badana zmienna	Ważność [%]
Prędkość pojazdu [km/h]	100
Względne obciążenie silnika [%]	87
Względne położenie przepustnicy [%]	65

Ponownie zmiennymi o największym wpływie okazały się zmienne *prędkość pojazdu* oraz *względne obciążenie silnika*, a najmniejszym *względne położenie przepustnicy*.

Opracowane powyżej modele zapisano w formie kodu PMML, a następnie wdrożono do przeprowadzenia badania w oparciu o zbiór danych testowych. W pierwszym kroku dokonano obliczenia przewidywanych wartości *emisji CO₂* odpowiednio dla modelu z algorytmem CART, lasu losowego oraz drzew wzmacnianych. Obliczone błędy predykcji przedstawiono w tabeli 36.

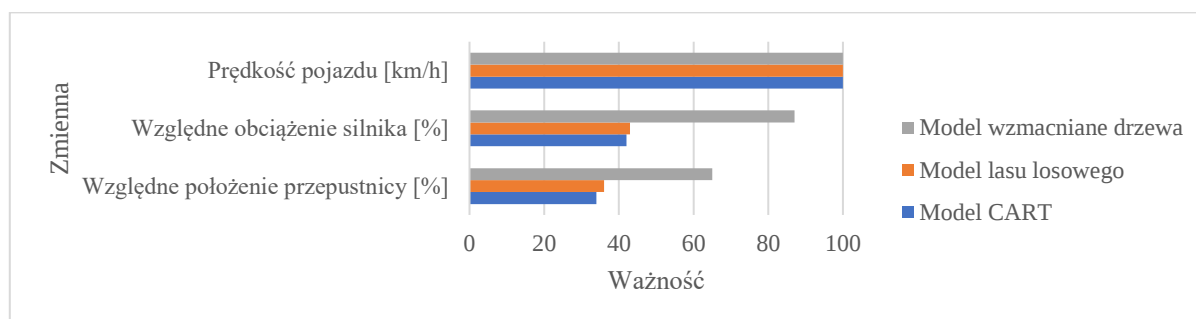
Tabela 36. Porównanie możliwości predykcyjnych modeli drzew regresyjnych w trybie ruszania pojazdu

Rodzaj błędu prognozy	Model drzewa decyzyjnego CART	Model lasu losowego	Model drzew wzmacnianych
Średnia kwadratów reszt	9 308,50	8 770,90	52 66,27
Średni błąd bezwzględny	69,92	68,00	42,04

Rodzaj błędu prognozy	Model drzewa decyzyjnego CART	Model lasu losowego	Model drzew wzmocnianych
Względny średni błąd kwadratowy	0,31	0,22	0,34
Względne odchylenie przeciętne	0,38	0,34	0,25
Średni bezwzględny błąd procentowy	0,44	0,42	0,22

Na podstawie powyższej tabeli stwierdzić należy, że najlepszymi zdolnościami predykcyjnymi cechuje się model drzew wzmocnianych, natomiast największe błędy prognozy uzyskano przy zastosowaniu modelu drzewa decyzyjnego w oparciu o algorytm CART.

Na rysunku 56 przedstawiono ranking predyktorów dla trzech omawianych modeli w odniesieniu do rozpatrywanego trybu ruszania pojazdu.



Rys. 56. Ranking predyktorów modeli w trybie ruszania pojazdu

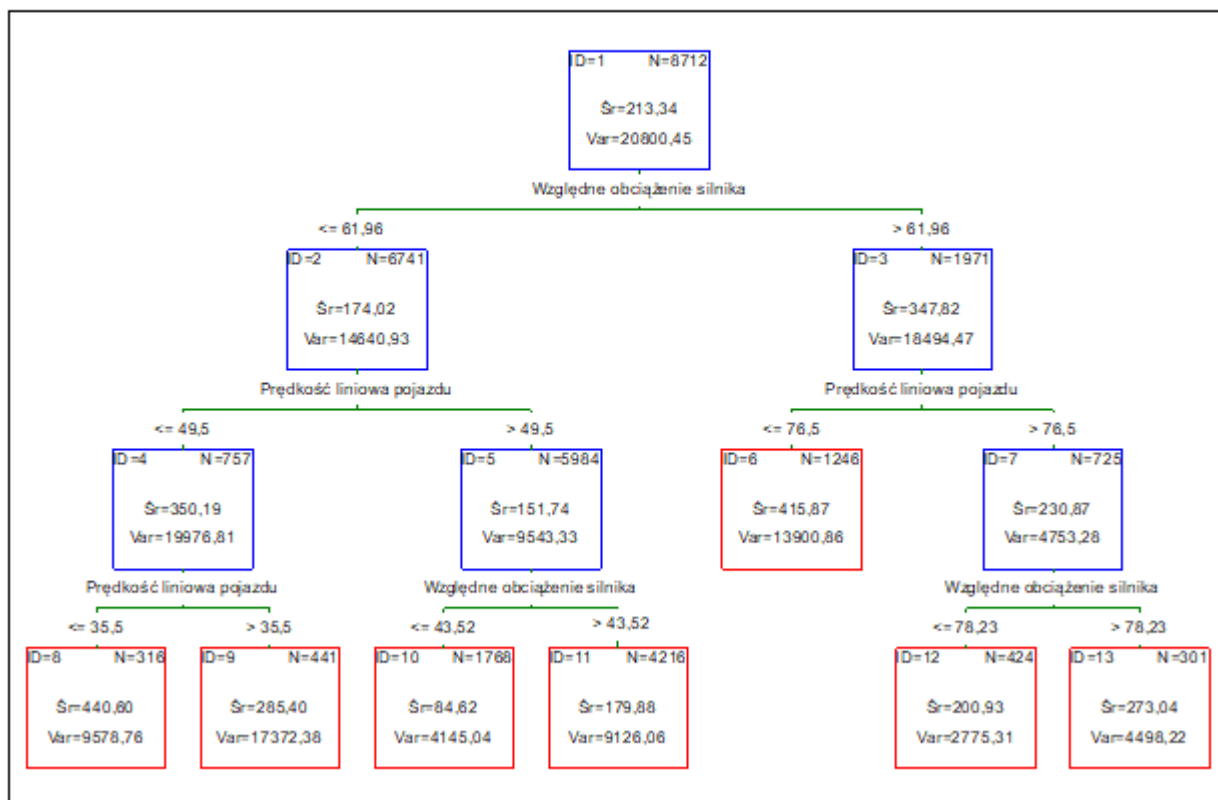
Czynnikiem mającym największy wpływ na poziom emisji CO₂ dla wszystkich utworzonych modeli jest *prędkość pojazdu*. Najmniejszy wpływ cechuje zmienną *względne położenie przepustnicy*. Zauważyć należy, że dla modeli lasu losowego oraz CART wartości w rankingu predyktorów są zbliżone dla wszystkich zmiennych. Natomiast w modelu drzew wzmocnianych zauważalne jest większy wpływ zmiennych *względne obciążenie silnika* oraz *względne położenie przepustnicy* na badane zjawisko, w odniesieniu do pozostałych modeli.

5.3.2. Modele regresyjnych drzew decyzyjnych dla jazdy w trybie miejskim

W kolejnej części pracy przystąpiono do budowy drzew decyzyjnych do modelowania *emisji CO₂* dla jazdy pojazdu w trybie miejskim. Jako pierwszy zbudowano model drzewa decyzyjnego CART.

Wynik analizy przedstawiono na rysunku 57.

Drzewo dla Emisja CO₂
Liczba węzłów dzielonych: 6, liczba węzłów końcowych: 7
Model: C&RT



Rys. 57. Drzewo decyzyjne wykorzystujące algorytm CART w trybie jazdy miejskiej

Na podstawie otrzymanych wyników, można wnioskować, że w przypadku jazdy przy *względным obciążeniu silnika* poniżej 43,52% oraz *prędkości pojazdu* powyżej 49,5 km/h poziom *emisji CO₂* jest najniższy i przyjmuje wartość średnią 84,62 g/km (liść nr 10). Natomiast, w przypadku jazdy z *prędkością pojazdu* poniżej 49,5 km/h oraz przy *względным obciążeniu silnika* poniżej 61,96% wartość średniej emisji CO₂ jest najwyższa i wynosi 440,60 g/km (liść 8).

W kolejnym kroku zbudowano ranking ważności predyktorów (tabela 37).

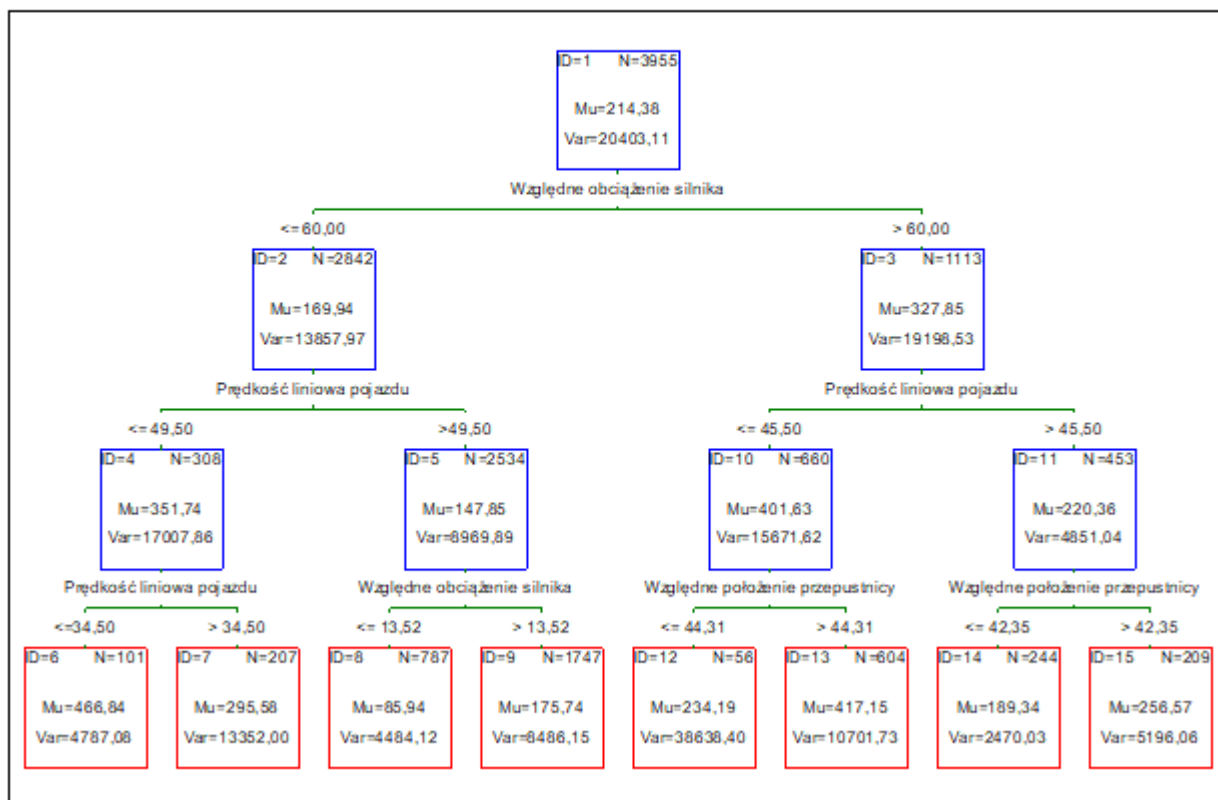
Tabela 37. Ranking predyktorów z zastosowaniem modelu CART w trybie jazdy miejskiej

Badana zmienna	Ważność [%]
Prędkość pojazdu [km/h]	100
Względne obciążenie silnika [%]	88
Względne położenie przepustnicy [%]	48

Zauważyć należy, że największy wpływ na badane zjawisko ma prędkość pojazdu, natomiast najmniejszy względne położenie przepustnicy.

Drugim z opracowywanych modeli był model lasu losowego. Wynik analizy (jednego wybranego drzewa) przedstawiono na rysunku 58.

Drzewo dla Emisja CO₂
Liczba węzłów dzielonych: 7, liczba węzłów końcowych: 8
Drzewo numer: 5



Rys. 58. Przykładowe drzewo decyzyjne z zastosowaniem lasu losowego w trybie jazdy miejskiej

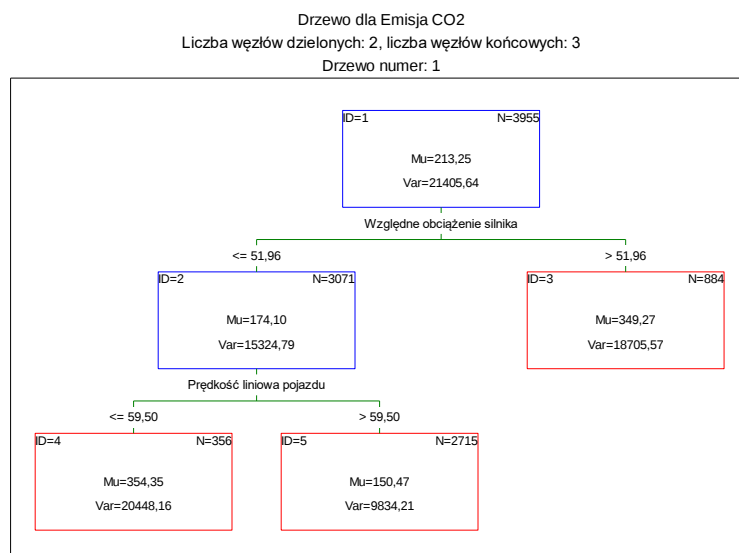
Następnie ponownie zbudowano ranking ważności predyktorów (tabela 38).

Tabela 38. Ranking predyktorów z zastosowaniem modelu lasu losowego w trybie jazdy miejskiej

Badana zmienna	Ważność [%]
Prędkość pojazdu [km/h]	100
Względne obciążenie silnika [%]	89
Względne położenie przepustnicy [%]	54

Ponownie największy wpływ miała zmienna *prędkość pojazdu*, najmniejszy zaś *względne położenie przepustnicy*.

Jako ostatni opracowano model wzmacnianych drzew. Wynik analizy przedstawiono na rysunku 59.



Rys. 59. Przykładowe drzewo decyzyjne z zastosowaniem drzew wzmacnianych w trybie miejskim

Zbudowano ranking ważności predyktorów (tabela 39).

Tabela 39. Ranking predyktorów w modelu z zastosowaniem drzew wzmacnianych w trybie jazdy miejskiej

Badana zmienna	Ważność [%]
Prędkość pojazdu [km/h]	100
Względne obciążenie silnika [%]	92
Względne położenie przepustnicy [%]	64

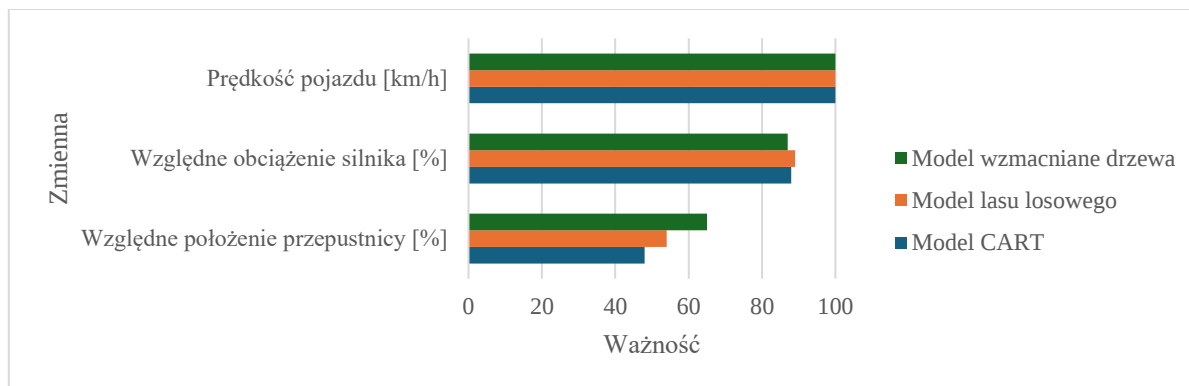
Zmiennymi o największym wpływie okazały się po raz kolejny zmienne *prędkość pojazdu* i *względne obciążenie silnika* a najmniejszym *względne położenie przepustnicy*. Ponownie powyższe modele zapisano w formie kodu PMML wykorzystując próbę testową, a następnie obliczono błędy predykcji *emisji CO₂* odpowiednio dla modelu z algorytmem CART, lasu losowego oraz drzew wzmacnianych (tabela 40).

Tabela 40. Porównanie możliwości predykcyjnych modeli drzew regresyjnych w trybie jazdy miejskiej

Rodzaj błędu prognozy	Model drzewa decyzyjnego CART	Model lasu losowego	Model drzew wzmacnianych
Średnia kwadratów reszt	6 237,00	4 970,85	5 735,22
Średni błąd bezwzględny	55,61	46,95	54,09
Względny średni błąd kwadratowy	0,05	0,04	0,05
Względne odchylenie przeciętne	0,16	0,13	0,15
Średni bezwzględny błąd procentowy	0,23	0,20	0,21

Na podstawie powyższej tabeli stwierdzić należy, że opracowane modele cechują się zbliżonymi zdolnościami predykcyjnymi, a najlepszym z nich jest model drzew losowych.

Na rysunku 60 przedstawiono ranking predyktorów dla trzech omawianych modeli w odniesieniu do rozpatrywanego trybu jazdy miejskiej.



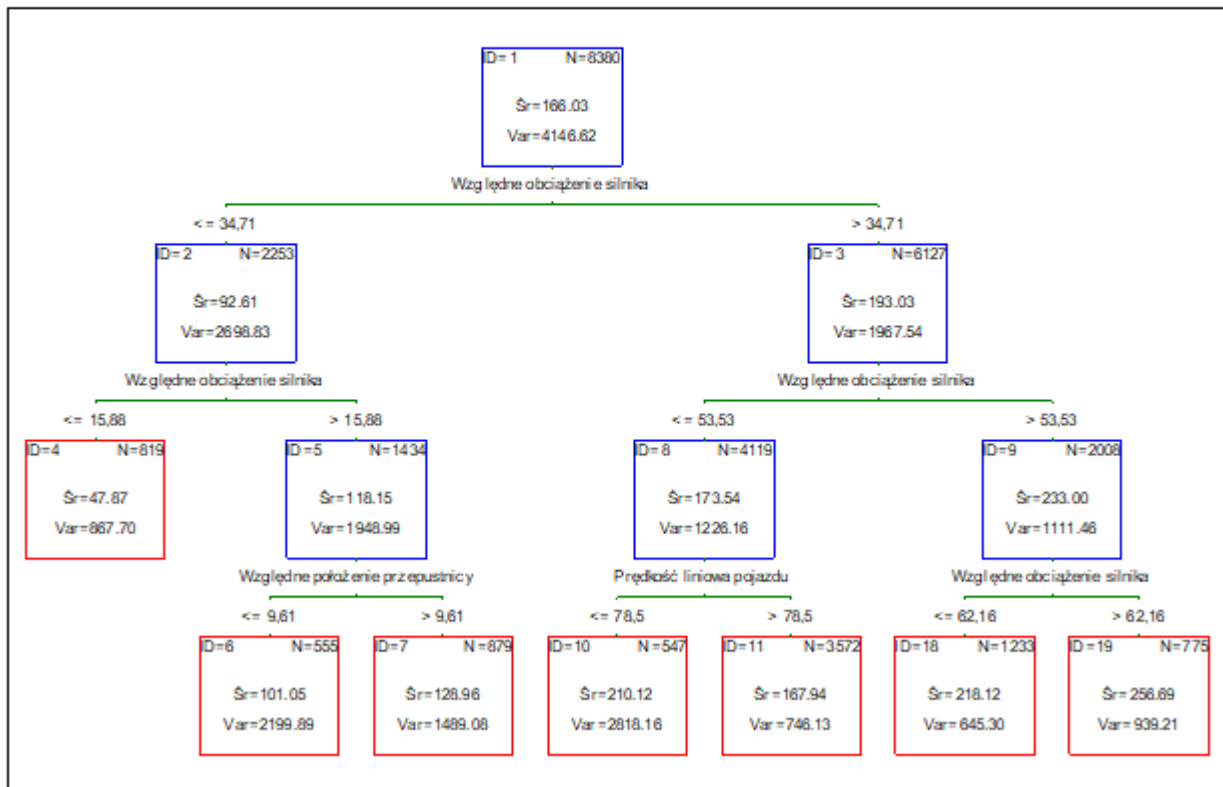
Rys. 60. Ranking predyktorów modeli w trybie jazdy miejskiej

Czynnikami mającymi największy wpływ na poziom *emisji CO₂* dla wszystkich utworzonych modeli są kolejno *prędkość pojazdu*, *względne obciążenie silnika* oraz *względne położenie przepustnicy*.

5.3.3. Modele regresyjnych drzew decyzyjnych podczas jazdy w trybie pozamiejskim

W kolejnej części pracy przystąpiono do modelowania i analizy wpływu wytypowanych czynników na poziom *emisji CO₂* podczas jazdy w trybie pozamiejskim. Ponownie rozpoczęto od modelu drzewa decyzyjnego z algorytmem CART. Wynik analizy przedstawiono na rysunku 61.

Drzewo dla Emisja CO₂
Liczba węzłów dzielonych: 6, liczba węzłów końcowych: 7
Model: C&RT



Rys. 61. Drzewo decyzyjne algorytmem CART w czasie jazdy w trybie pozamiejskim

Na podstawie otrzymanych wyników, można wnioskować, że w przypadku jazdy przy *względny obciążeniu silnika* poniżej 15,88% poziom emisji CO₂ jest najniższy i przyjmuje wartość średnią 47,87 g/km (liść nr 4). Natomiast przy *względny obciążeniu silnika* powyżej 62,16% wartość średniej emisji CO₂ jest najwyższa i wynosi 256,59 g/km.

W kolejnym kroku zbudowano ranking ważności predyktorów (tabeli 41).

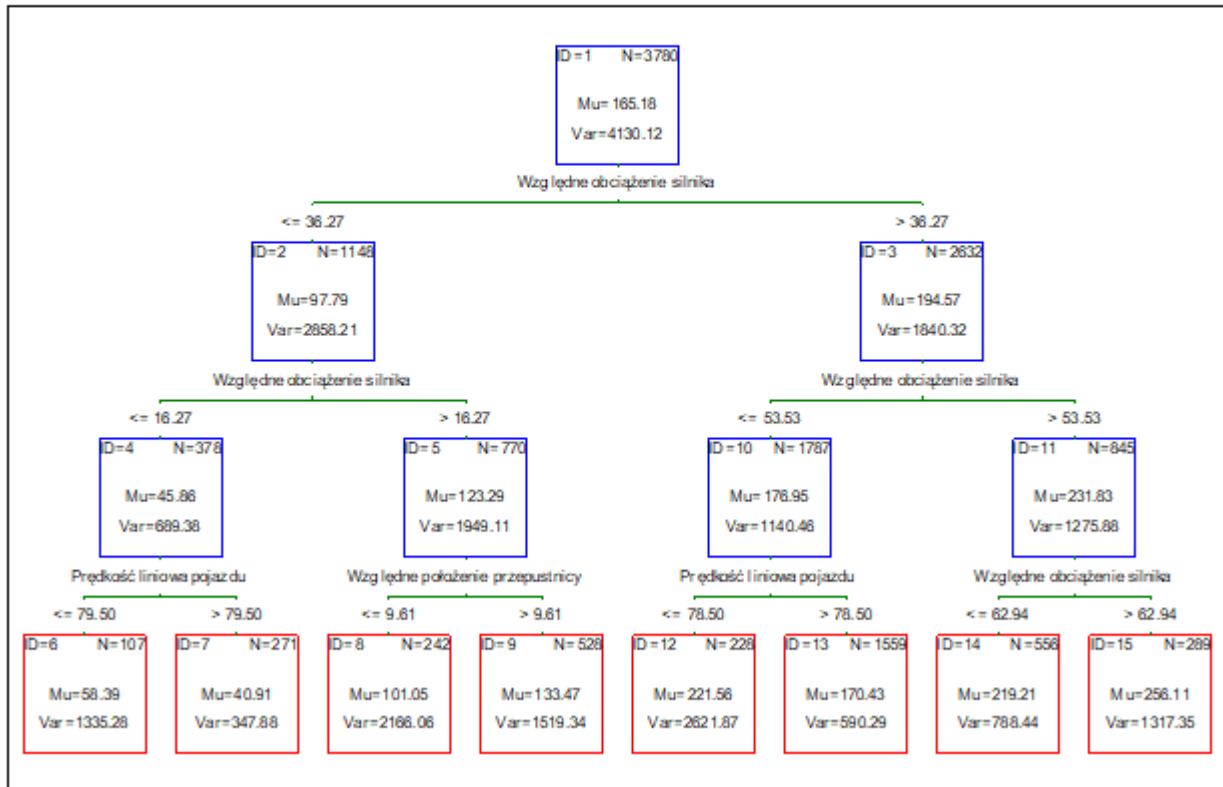
Tabela 41. Ranking predyktorów dla modelu CART w czasie jazdy w trybie pozamiejskim

Badana zmienna	Ważność [%]
Względne obciążenie silnika [%]	100
Względne położenie przepustnicy [%]	92
Prędkość pojazdu [km/h]	15

Zauważyć należy, że największy wpływ na badane zjawisko ma *względne obciążenie silnika* oraz *względne położenie przepustnicy*, natomiast najmniejszy *prędkość pojazdu*.

Drugim z opracowywanych modeli był model lasu losowego. Wynik analizy (dla wybranego drzewa) przedstawiono na rysunku 62.

Drzewo dla Emisja CO₂
Liczba węzłów dzielonych: 7, liczba węzłów końcowych: 8
Drzewo numer: 1



Rys. 62. Przykładowe drzewo decyzyjne z zastosowaniem lasu losowego w trybie jazdy pozamiejscowej

W kolejnym kroku zbudowano ranking ważności predyktorów (tabela 42).

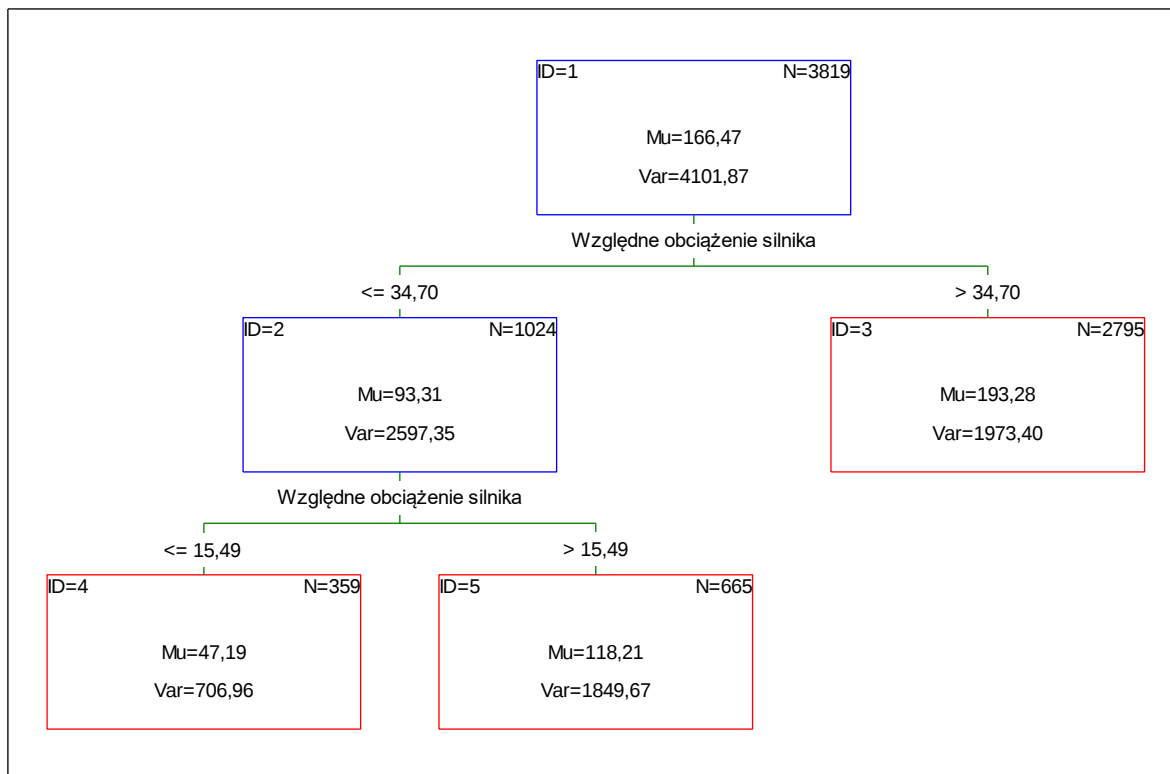
Tabela 42. Ranking predyktorów z zastosowaniem modelu lasu losowego w czasie jazdy w trybie pozamiejscowym

Badana zmienna	Ważność [%]
Względne obciążenie silnika [%]	100
Względne położenie przepustnicy [%]	92
Prędkość pojazdu [km/h]	20

Ponownie największy wpływ miały *względne obciążenie silnika* oraz *względne położenie przepustnicy*, najmniejszy zaś *prędkość pojazdu*.

Jako ostatni opracowano model wzmacnianych drzew. Wynik analizy przedstawiono na rysunku 63.

Drzewo dla Emisja CO₂
Liczba węzłów dzielonych: 2, liczba węzłów końcowych: 3
Drzewo numer: 1



Rys. 63. Przykładowe drzewo decyzyjne z zastosowaniem drzew wzmacnianych w trybie jazdy pozamiejskiej

Na tej podstawie zbudowano ranking ważności predyktorów (tabela 43).

Tabela 43. Ranking predyktorów w modelu z zastosowaniem drzew wzmacnianych dla jazdy w trybie pozamiejskim

Badana zmienna	Ważność [%]
Względne obciążenie silnika [%]	100
Względne położenie przepustnicy [%]	90
Prędkość pojazdu [km/h]	52

Ponownie największy wpływ miały *względne obciążenie silnika* oraz *względne położenie przepustnicy*, najmniejszy zaś *prędkość pojazdu*.

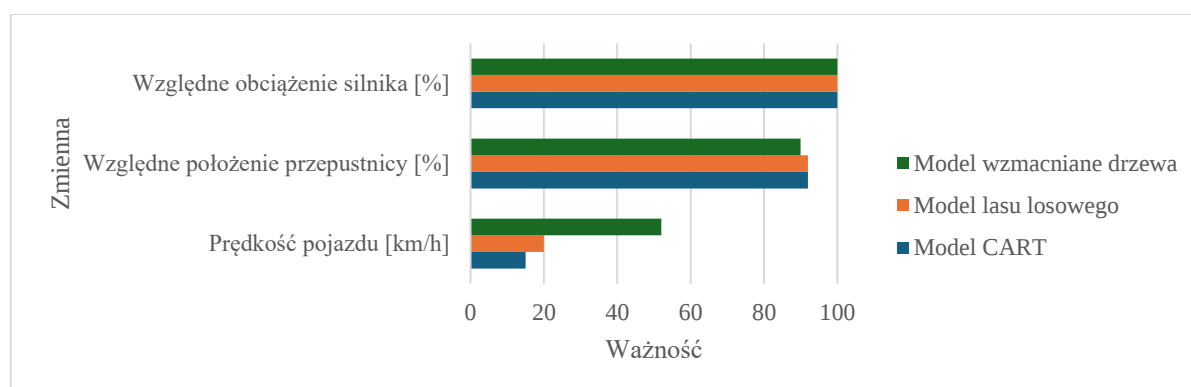
Opracowane powyżej modele analogicznie zapisano w formie kodu PMML, a następnie przeprowadzono badania w oparciu o próbę testową. Dokonano obliczenia przewidywanych wartości *emisji CO₂* odpowiednio dla modelu z algorytmem CART, lasu losowego oraz drzew wzmacnianych i na tej podstawie obliczono błędy predykcji, które przedstawiono w tabeli 44.

Tabela 44. Porównanie możliwości predykcyjnych modeli drzew regresyjnych w czasie jazdy w trybie pozamiejskim

Rodzaj błędu prognozy	Model drzewa decyzyjnego CART	Model lasu losowego	Model drzew wzmacnianych
Średnia kwadratów reszt	913,47	737,16	802,75
Średni błąd bezwzględny	17,99	14,93	16,95
Względny średni błąd kwadratowy	0,07	0,06	0,34
Względne odchylenie przeciętne	0,14	0,12	0,15
Średni bezwzględny błąd procentowy	0,19	0,16	0,18

Na podstawie powyższej tabeli stwierdzić należy, że najlepszymi zdolnościami predykcyjnymi cechuje się model lasu losowego, natomiast największe błędy prognozy uzyskano przy zastosowaniu modelu drzewa decyzyjnego w oparciu o algorytm CART.

Na rysunku 64 przedstawiono ranking predyktorów dla trzech omawianych modeli w odniesieniu do rozpatrywanego trybu jazdy pozamiejskiej.



Rys. 64. Ranking predyktorów modeli w trybie jazdy pozamiejskiej

Czynnikami mającymi największy wpływ na poziom emisji CO_2 dla wszystkich utworzonych modeli są kolejno *względne obciążenie silnika*, *względne położenie przepustnicy* oraz *prędkość pojazdu*. Ponadto modele lasu losowego oraz CART osiągają bardzo zbliżone wartości w rankingu predyktorów dla zmiennych *względne obciążenie silnika* oraz *względne położenie przepustnicy*. Przeciwna sytuacja ma miejsce w przypadku zmiennej *prędkość pojazdu*, której ważność w modelu drzew wzmacnianych jest wyższa niż w pozostałych modelach.

5.3.4. Porównanie modeli w poszczególnych trybach jazdy

W ostatnim kroku dokonano porównania wyników otrzymanych dla poszczególnych modeli w zależności od trybu jazdy pojazdu (w trybie ruszania pojazdu, tryb miejski, tryb pozamiejski). W tabeli 45 zaprezentowano zestawienie rankingu predyktorów w trzech

wariantach w celu uzyskania odpowiedzi, która ze zmiennych objaśniających ma największy wpływ na modelowane zjawisko tj. *emisję CO₂*.

Tabela 45. Porównanie ważności predyktorów dla poszczególnych modeli drzew regresyjnych w różnych trybach jazdy

Tryb jazdy	Model drzewa decyzyjnego CART		Model lasu losowego		Model drzew wzmacnianych	
	Zmienna objaśniająca	Ważność [%]	Zmienna objaśniająca	Ważność [%]	Zmienna objaśniająca	Ważność [%]
Tryb ruszanie pojazdu	Prędkość pojazdu [km/h]	100	Prędkość pojazdu [km/h]	100	Prędkość pojazdu [km/h]	100
	Względne obciążenie silnika [%]	42	Względne obciążenie silnika [%]	43	Względne obciążenie silnika [%]	87
	Względne położenie przepustnicy [%]	34	Względne położenie przepustnicy [%]	36	Względne położenie przepustnicy [%]	65
Tryb jazdy miejskiej	Prędkość pojazdu [km/h]	100	Prędkość pojazdu [km/h]	100	Prędkość pojazdu [km/h]	100
	Względne obciążenie silnika [%]	88	Względne obciążenie silnika [%]	89	Względne obciążenie silnika [%]	92
	Względne położenie przepustnicy [%]	48	Względne położenie przepustnicy [%]	54	Względne położenie przepustnicy [%]	64
Tryb jazdy pozamiejskiej	Względne obciążenie silnika [%]	100	Względne obciążenie silnika [%]	100	Względne obciążenie silnika [%]	100
	Względne położenie przepustnicy [%]	92	Względne położenie przepustnicy [%]	92	Względne położenie przepustnicy [%]	90
	Prędkość pojazdu [km/h]	15	Prędkość pojazdu [km/h]	20	Prędkość pojazdu [km/h]	52

W przypadku modeli zbudowanych dla momentu ruszania pojazdu oraz jazdy w trybie miejskim, czynnikiem mającym największy wpływ na poziom *emisji CO₂* jest *prędkość pojazdu*. W przypadku jazdy w trybie pozamiejskim, czynnikami o największym wpływie na poziom *emisji CO₂* są zmienne *względne obciążenie silnika* oraz *względne położenie przepustnicy*.

Tabela 46. Porównanie zdolności predykcyjnych poszczególnych modeli drzew regresyjnych w różnych trybach jazdy

Tryb jazdy	Rodzaj błędu prognozy	Model drzewa decyzyjnego CART	Model lasu losowego	Model drzew wzmacnianych
Tryb ruszanie pojazdu	Średnia kwadratów reszt	9 308,50	8 770,90	5 266,27
	Średni błąd bezwzględny	69,92	68,00	42,04
	Względny średni błąd kwadratowy	0,31	0,22	0,34
	Względne odchylenie przeciętne	0,38	0,34	0,25
	Średni bezwzględny błąd procentowy	0,44	0,42	0,22
Tryb jazdy miejskiej	Średnia kwadratów reszt	6 237,00	4 970,85	5 735,22
	Średni błąd bezwzględny	55,61	46,95	54,09
	Względny średni błąd kwadratowy	0,05	0,04	0,05
	Względne odchylenie przeciętne	0,16	0,13	0,15
	Średni bezwzględny błąd procentowy	0,23	0,20	0,21
Tryb jazdy pozamiejskiej	Średnia kwadratów reszt	913,47	737,16	802,75
	Średni błąd bezwzględny	17,99	14,93	16,95
	Względny średni błąd kwadratowy	0,07	0,06	0,34
	Względne odchylenie przeciętne	0,14	0,12	0,15
	Średni bezwzględny błąd procentowy	0,19	0,16	0,18

Na podstawie oszacowanych błędów prognozy zauważyć należy, że najlepszym dopasowaniem do danych empirycznych cechuje się model oszacowany dla jazdy w trybie pozamiejskim (tabela 46). Średni bezwzględny błąd prognozy kształtuje się na poziomie 14,93 g/km i 17,99 g/km (odpowiednio dla modelu lasu losowego oraz z algorytmem CART). Modele cechuje również zadowalający poziom średniego bezwzględnego błędu procentowego

(MAPE), kształtującego się na poziomie 16%, 18% i 19%, odpowiednio dla modelu lasu losowego, drzew wzmacnianych i drzew decyzyjnych z algorytmem CART.

Najślabsze dopasowanie do danych empirycznych osiągnięto dla modelu w trybie ruszania pojazdu, dla którego średni błąd bezwzględny prognozy wynosi 42 g/km, 68 g/km i 69 g/km odpowiednio dla modeli CART, lasu losowego i drzew wzmacnianych. Natomiast średni bezwzględny błąd procentowy kształtuje się na poziomie 44% dla modelu z algorytmem CART oraz 22% dla drzew wzmacnianych (poziom zadowalający).

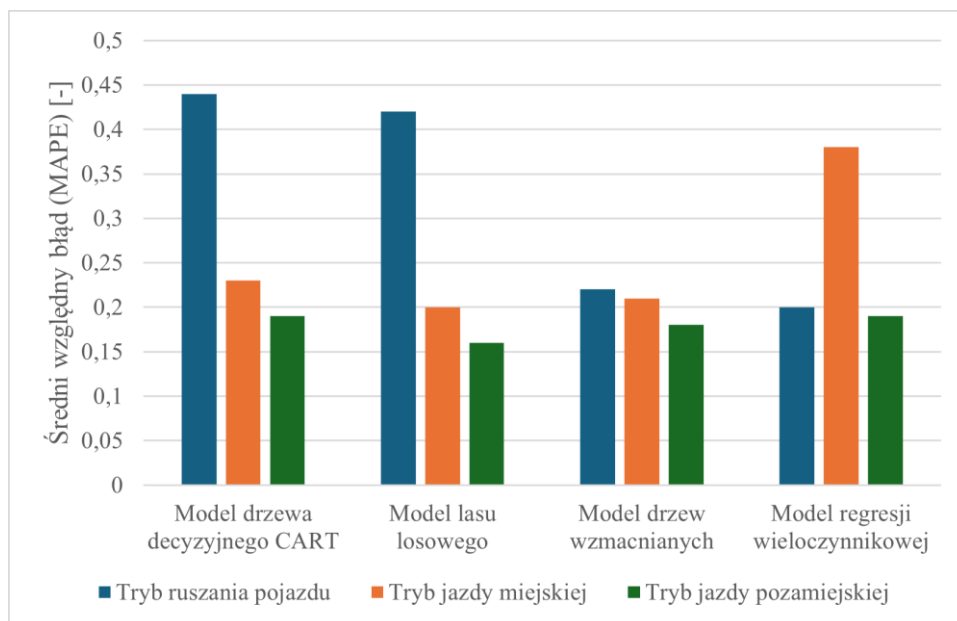
W przypadku modeli *emisji CO₂* podczas jazdy w trybie miejskim średni błąd bezwzględny wynosi 46 g/km (model lasu losowego), 54 g/km (model drzew wzmacnianych) oraz 55 g/km (model z algorytmem CART). Średni bezwzględny błąd procentowy (MAPE) dla powyższej jazdy wynosi odpowiednio 20% dla modelu lasu losowego oraz 23% dla modelu z algorytmem CART (poziom zadowalający).

6. ZAKOŃCZENIE

6.1. Podsumowanie przeprowadzonych badań

Problematyka emisji CO₂ stanowi niezwykle ważny i aktualny problem w obecnych czasach. Wysiłki podejmowane przez UE stanowią bezpośrednie potwierdzenie skali tego zjawiska. Strategia UE dotycząca zmniejszenia emisji CO₂ z pojazdów osobowych opiera się na trzech filarach, które stanowią: dobrowolne porozumienie z producentami nowo produkowanych pojazdów, dyrektywa dotycząca oznakowania pojazdów pod kątem emisji CO₂ oraz propozycja harmonizacji instrumentów fiskalnych dotyczących samochodów osobowych w krajach członkowskich UE. Z kolei jednym z najważniejszych działań dotyczących sektora transportu drogowego jest zachęcanie do ekonomicznej i ekologicznej jazdy kierowców. W związku z tym wszystkie inicjatywy zmniejszające emisję CO₂ są niezwykle pożądane, zarówno z perspektywy ochrony środowiska, jak i efektywności energetycznej. Istotną rolę pełni również świadomość kierowców i potencjał kształtowania proekologicznych nawyków, co w dłuższej perspektywie sprzyja osiągnięciu celów klimatycznych i zrównoważonego rozwoju i dlatego stało się genezą niniejszej rozprawy.

W kontekście omawianej problematyki, w celu minimalizacji emisji CO₂, kluczowe znaczenie mają monitorowanie oraz analiza emisji CO₂ w pojazdach osobowych. Zastosowanie interfejsów diagnostycznych OBD umożliwia pozyskanie szczegółowych informacji dotyczących parametrów pracy silnika pojazdu osobowego oraz wartości emisji CO₂ w czasie rzeczywistym. Stanowi to fundament prowadzonych w dysertacji rozważań. Z kolei zastosowanie modelowania matematycznego, a dokładnie regresji wieloczynnikowej oraz regresyjnych drzew decyzyjnych pozwoliło na udzielenie odpowiedzi na przyjęty problem badawczy. O skuteczności zaproponowanych modeli dowodzą otrzymane wyniki badań, co zobrazowano na przykładzie średniego bezwzględnego błędu procentowego (MAPE), którego wartości dla wszystkich utworzonych modeli drzew decyzyjnych oraz regresji wieloczynnikowej w zależności od trybu jazdy przedstawiono zbiorczo na rysunku 65.



Rys. 65. Średni bezwzględny błąd procentowy (MAPE) dla modeli drzew decyzyjnych oraz regresji wieloczynnikowej w zależności od trybu jazdy

Średni bezwzględny błąd procentowy (MAPE) w przypadku trybu ruszania pojazdu jest najmniejszy dla modelu regresji wieloczynnikowej, natomiast w przypadku trybów jazdy miejskiej oraz pozamiejskiej dla modelu lasu losowego.

Opracowane modele pozwoliły również na szczegółowe konkluzje. W przypadku modelu regresji wieloczynnikowej stwierdzono, iż:

- Niezależnie od trybu jazdy, wzrost *prędkości pojazdu* o 1 km/h powoduje zmniejszenie poziomu *emisji CO₂*.
- Wzrost o 1 % *względnego obciążenia silnika* powoduje wzrost *emisji CO₂*. Największą wartość wzrostu osiągnięto w trybie miejskim 4,23 g/km.
- Wzrost o 1% *względnego położenia przepustnicy* powoduje w przypadku trybu ruszania pojazdu spadek *emisji CO₂* o 0,75 g/km, natomiast w pozostałych trybach wzrost (w trybie miejskim aż o 9,02 g/km).

Badania wykazały, że największy średni poziom *emisji CO₂* do środowiska ma miejsce podczas ruszania pojazdu (365,60 g/km), natomiast najmniejszy w czasie poruszania się pojazdu w trybie pozamiejskim (166,18 g/km). Niezależnie od trybu jazdy wzrost prędkości pojazdu powoduje zmniejszenie *emisji CO₂* do środowiska. W czasie ruszania pojazdu z każdym wzrostem prędkości pojazdu o 1 km/h, *emisja CO₂* spada o 2,59 g/km, podczas jazdy w trybie miejskim spada o 4,11 g/km, natomiast w czasie jazdy pozamiejskiej wraz ze wzrostem emisja spada o 0,56 g/km. Z kolei w przypadku wzrostu o 1% *względnego obciążenia silnika*, następuje wzrost *emisji CO₂* o 3,06 g/km, 4,23 g/km oraz 3,21 g/km (podczas ruszania pojazdu, w trybie miejskim oraz podczas jazdy w trybie pozamiejskim). Natomiast zmiana

względne położenie przepustnicy o 1% powoduje odpowiednio spadek emisji CO_2 o 0,75 g/km (podczas ruszania pojazdu) oraz wzrost o 9,02 g/km i 2,34 g/km (odpowiednio tryb jazdy miejskiej, tryb jazdy pozamiejskiej). Ponadto analiza poszczególnych modeli wykazała, że największym błędem estymacji cechuje się model regresji w trybie jazdy miejskiej (88,02 g/km), natomiast najmniejszym model regresji w czasie jazdy pozamiejskiej (30,19 g/km). Najmniejszym średnim błędem bezwzględnym cechuje się natomiast model dla w trybie ruszania pojazdu (37,79 g/km), natomiast największym model jazdy w trybie miejskim (60,57 g/km).

W przypadku modeli drzew decyzyjnych stwierdzono, iż:

- W trybach jazdy: ruszanie pojazdu oraz miejskim, największy wpływ na poziom *emisji CO_2* ma *prędkość pojazdu*. W trybie pozamiejskim: względne obciążenie silnika oraz względne położenie przepustnicy.
- W trybie ruszania pojazdu najniższa wartość *emisji CO_2* (233,22 g/km) ma miejsce w przypadku, gdy *prędkość pojazdu* jest powyżej 19,5 km/h oraz *względne obciążenie silnika* poniżej 66,86%. Najwyższą (484,91 g/km) natomiast w przypadku *prędkości pojazdu* poniżej 6,5 km/h i *względnym obciążeniu silnika* powyżej 42,55%.
- W trybie jazdy miejskiej najniższą wartość *emisji CO_2* (84,62 g/km) otrzymano dla *względnego obciążenia silnika* poniżej 42,52% oraz *prędkości pojazdu* powyżej 49,5 km/h. Najwyższą natomiast przy *względnym obciążeniu silnika* poniżej 61,96 i *prędkości pojazdu* poniżej 49,5 km/h.
- W trybie jazdy pozamiejskiej, przy *względnym obciążeniu silnika* poniżej 15,88%, poziom *emisji CO_2* jest najniższy (47,87 g/km). Natomiast najwyższy (256,59 g/km) w przypadku *względnego obciążenia silnika* powyżej 62,16%.

W przypadku modeli utworzonych w trybie ruszania pojazdu oraz w trybie jazdy miejskiej, czynnikiem mającym największy wpływ na poziom *emisji CO_2* jest *prędkość pojazdu*. W czasie ruszania pojazdu podobny wpływ na *emisję CO_2* mają *względne obciążenie silnika* oraz *względne położenie przepustnicy* dla modeli CART (42/100, 34/100) oraz lasu losowego (43/100, 36/100) z wyjątkiem modelu wzmacnianych drzew (87/100, 65/100). W przypadku modelu podczas jazdy w trybie miejskim kolejność ważności zmiennych jest identyczna. Natomiast w przypadku jazdy pozamiejskiej, czynnikami o największym wpływie na poziom *emisji CO_2* są zmienne *względne obciążenie silnika* oraz *względne położenie przepustnicy*. Na podstawie oszacowanych błędów prognozy zauważyć należy, że największą dobrocią dopasowania cechuje się model oszacowany dla jazdy pozamiejskiej. Średni bezwzględny błąd prognozy kształtuje się na poziomie 14,93 g/km oraz 17,99 g/km (kolejno

dla modelu lasu losowego oraz z algorytmem CART). Modele też cechuje także poziom średniego bezwzględnego błędu procentowego (MAPE) kształtujący się na poziomie od 16%, 18% i 19%. Najniższą jakość dopasowania osiągnięto dla modelu w czasie ruszania pojazdu, dla którego średni błąd bezwzględny prognozy wynosi od 42 g/km oraz 69 g/km odpowiednio dla modeli drzew wzmacnianych oraz modelu z algorytmem CART. Natomiast średni bezwzględny błąd procentowy kształtuje się na poziomie 44% dla modelu z algorytmem CART oraz 22% dla drzew wzmacnianych. W przypadku modeli zbudowanych dla jazdy w trybie miejskim średni błąd bezwzględny wynosi 46 g/km (model lasu losowego) oraz 55 g/km (model z algorytmem CART). Średni bezwzględny błąd procentowy (MAPE) dla powyższej jazdy wynosi odpowiednio 20% dla modelu lasu losowego oraz 23% dla modelu z algorytmem CART.

6.2. Wnioski

Rozwiązaniem postawionego w pracy problemu badawczego jest odpowiedź na pytanie, w jaki sposób można wpływać na emisję dwutlenku węgla z pojazdów osobowych na poziomie pojedynczego użytkownika. Przeprowadzone badania wykazały, iż możliwe jest opracowanie modelu oceny kształtowania poziomu emisji CO₂ na podstawie informacji pozyskanych z interfejsów diagnostycznych oraz wykorzystanie otrzymanych wyników do zmniejszenia emisji CO₂ poprzez kształtowanie stylu jazdy kierowców. Analiza danych z interfejsów diagnostycznych pozwoliła na identyfikację i pozyskanie kluczowych parametrów wpływających na emisję CO₂, a przeprowadzone w rozprawie badania wykazały, które z nich są najważniejsze w podejmowaniu bardziej ekologicznych decyzji. Tym samym, prezentowane badania stanowią istotny krok w kierunku redukcji emisji gazów cieplarnianych w sektorze transportu drogowego. Powyższe dowodzi także, że cel pracy został osiągnięty a teza badawcza zweryfikowana pozytywnie.

Zastosowanie metod regresji wieloczynnikowej oraz regresyjnych drzew decyzyjnych w celu modelowania emisji CO₂ pozwoliło nie tylko na zrozumienie ogólnych zależności, opisujących emisję CO₂, ale także na wskazanie konkretnych czynników, których kontrola może przyczynić się do jej redukcji. W pracy wykazano, iż monitorowanie podczas jazdy *prędkości pojazdu, względnego obciążenia silnika oraz względnego położenia przepustnicy* pozwala na tę kontrolę, dzięki czemu pojedynczy kierowca pojazdu osobowego, niezależnie od trybu poruszania się pojazdu, jest w stanie wpływać na emisję CO₂. Tym samym wykazano, iż wymienione powyżej, mierzalne i sterowalne na poziomie użytkownika pojazdu osobowego

czynniki, mają istotny wpływ na poziom emisji CO₂, co stanowiło pierwszy cel szczegółowy pracy. Warto również podkreślić, że w odróżnieniu od standardowych testów laboratoryjnych, które często bazują na uśrednionych parametrach i założonych warunkach jazdy, analiza danych z OBD pozwala uwzględnić dynamiczne zmiany wynikające z rzeczywistego użytkowania pojazdu, tj. zmiany prędkości, obciążenia silnika, prędkości obrotowej silnika, momentu obrotowego, mocy silnika czy względnego położenia przepustnicy.

Ponadto w wyniku przeprowadzonego badania stwierdzono, że interfejs OBD, który jest powszechnie stosowany w pojazdach osobowych, może stanowić doskonałe narzędzie ograniczania negatywnego wpływu na środowisko, dostępne dla pojedynczego użytkownika, bez dodatkowych kosztów. Zapewnienie ciągłego dostępu do rejestru wybranych parametrów pracy silnika oraz innych, kluczowych charakterystyk pracy pojazdu pozwala na podejmowanie świadomych decyzji dotyczących np. prędkości pojazdu. Pozwoliło to osiągnąć drugi cel szczegółowy pracy.

W toku prowadzonych rozważań stwierdzono również, że zwiększanie *względnego obciążenia silnika* prowadzi do wzrostu *emisji CO₂*. Wyniki wykazały, że utrzymywanie tego parametru na poziomie nie większym niż 45%-50% może pozytywnie wpływać na redukcję CO₂. Z kolei analiza wyników badań *względnego położenia przepustnicy* dostarczyła istotnych informacji na temat wpływu sposobu operowania pedałem przyspieszenia na *emisję CO₂*. Otwarcie przepustnicy powyżej 50% skutkuje wzrostem *emisji CO₂*, natomiast płynne przyspieszanie pozwala na zmniejszanie poziomu *emisji CO₂*. Badania nad wpływem *prędkości pojazdu* na *emisję CO₂* wykazały, że najbardziej efektywna pod względem zużycia paliwa i emisji CO₂ jest jazda ze stałą, umiarkowaną prędkością. Częste zmiany prędkości, nagłe przyspieszanie i hamowanie prowadzą do większego obciążenia silnika oraz wzrostu emisji. W pracy wykazano, iż podczas ruszania pojazdu, korzystnie jest powoli zwiększać prędkość pojazdu (minimum lokalne emisji CO₂ dla ok. 15 km/h, rys. 35) zwracając uwagę, aby nie obciążać zbyt mocno silnika (umiarkowane względne obciążenie silnika na poziomie 20-60%, rys. 37). Ponadto najlepszą strategią podczas jazdy miejskiej okazało się poruszanie się z możliwie najwyższą dopuszczalną w tym obszarze prędkością (50-90 km/h, rys. 35), dążąc do minimalizacji względnych wartości zarówno położenia przepustnicy jak i obciążenia silnika. Jednym ze sposobów na realizację tego jest unikanie gwałtownych zmian prędkości i obciążenia silnika, a podczas jazdy pozamiejskiej utrzymywanie stałej prędkości pojazdu (90-130 km/h). Od prędkości wynoszącej 140 km/h, następuje nagły wzrost emisji (rys. 35). Stosując wymienione rekomendacje pojedynczy użytkownik ruchu poprzez swoją ekologiczną jazdę wpływa na redukcję emisji gazów cieplarnianych, wpisując się w strategię

zrównoważonego rozwoju tj. ekologicznych nawyków jazdy, optymalizację użytkowania pojazdów oraz wdrażanie nowych technologii. W ten sposób zrealizowano ostatni, trzeci cel szczegółowy rozprawy i jednocześnie potwierdzono użyteczny charakter pracy.

Z perspektywy badawczej otrzymane wyniki mają istotne znaczenie naukowe, ponieważ pokazują w jaki sposób dane pozyskane bezpośrednio z systemu OBD mogą być wykorzystywane do modelowania emisji CO₂ w rzeczywistych warunkach eksploatacyjnych pojazdów. Zaproponowane rozwiązanie postawionego problemu badawczego ma również duże znaczenie dla rozwoju teorii modelowania matematycznego. Pod względem metodologicznym przeprowadzone badania otwierają nowe możliwości w zakresie precyzyjnego modelowania substancji szkodliwych i CO₂ na poziomie pojedynczego użytkownika. Wykorzystanie metod statystycznych umożliwiających identyfikację kluczowych czynników wpływających na poziom emisji, a następnie analizy regresyjnej oraz algorytmów uczenia maszynowego pozwoliło na budowę zaawansowanych modeli identyfikujących dane dotyczące emisji CO₂, a także na prognozowanie w tym zakresie. Wykorzystanie różnych metod modelowania i wykazanie, że takie badania na poziomie pojedynczego użytkownika są w ogóle możliwe pozwoliło stwierdzić, że precyzyjna analiza emisji CO₂ w oparciu o dane z systemu OBD może dostarczać wartościowych informacji na temat wpływu stylu jazdy na poziom emitowanego CO₂. Zastosowanie różnych metod modelowania pozwoliło na uzyskanie kompleksowego obrazu zależności między wybranymi czynnikami a emisją CO₂. Każda z tych technik wnosi unikalne korzyści do procesu analizy. Metody regresyjne umożliwiły precyzyjne określenie wpływu poszczególnych zmiennych na poziom emisji, co pozwoliło na ilościową ocenę zależności między parametrami jazdy (względne obciążenie silnika, względne położenie przepustnicy, prędkość pojazdu), a generowanym poziomem emisji CO₂. Pozwoliło to określić, które czynniki mają największy wpływ na emisję oraz w jakim stopniu ich zmiana może prowadzić do redukcji CO₂. Z kolei drzewa decyzyjne jako metoda nieliniowa, pozwoliły na identyfikację wzorców i reguł rządzących emisją CO₂ w różnych warunkach eksploatacyjnych pojazdu. Dzięki tej technice możliwe było wskazanie konkretnych sytuacji, w których emisja wzrasta – na przykład przy procentowym wzroście obciążenia silnika w trybie jazdy pozamiejskiej.

Zastosowanie regresji wieloczynnikowej oraz regresyjnych drzew decyzyjnych w modelowaniu emisji CO₂ pozwala na lepsze zrozumienie mechanizmów kształtowania się emisji CO₂. Utworzone modele można również zastosować do prognozowania.

Na podstawie przeprowadzonej analizy literatury w obszarze stylu jazdy kierowców stwierdzono, iż kierowca swoim zachowaniem podczas prowadzenia pojazdu może przyczynić

się do redukcji śladu węglowego związanego z transportem drogowym, dlatego kluczowe jest promowanie i wdrażanie technik ekologicznej jazdy.

Otrzymane wyniki przekładają się na konkretne zalecenia dla kierowców oraz sprzyjają klasyfikacji stylu jazdy pod kątem jej wpływu na emisję CO₂ i mogą być wykorzystane indywidualnie, w odniesieniu do pojedynczego kierowcy, co było głównym założeniem rozprawy. Jednak ich zastosowanie może być znacznie szersze. Mogą również stanowić podstawę do implementacji kompleksowych systemów wspomagania jazdy w pojazdach, wspierających kierowców w dostosowaniu stylu jazdy do osiągania bardziej ekologicznych przebiegów. Możliwość wykorzystania opracowanych modeli do monitorowania i analizy badanych parametrów w czasie rzeczywistym otwiera nowe perspektywy dla rozwoju inteligentnych systemów zarządzania emisją CO₂ w pojazdach, a także do opracowania skuteczniejszych strategii redukcji emisji w transporcie drogowym. Możliwość personalizacji zaleceń dla kierowców na podstawie indywidualnych odczytów otwiera nowe perspektywy w zakresie ekologicznej jazdy i ograniczania zużycia paliwa. Rezultaty pracy mogą również posłużyć do stworzenia innowacyjnych informatycznych narzędzi takich jak aplikacje mobilne wspierające kierowców w bardziej ekologicznej jeździe.

Wyniki i wnioski z przeprowadzonych badań ukazują duży potencjał w zakresie prowadzenia dalszych prac badawczych stanowiących rozszerzenie analiz o inne, nieuwzględnione w rozprawie czynniki wpływające na emisję CO₂, np. temperatura zewnętrzna, wysokość i nachylenie terenu itd., a także możliwość zastosowania metod sztucznej inteligencji w celu modelowania poziomu emisji CO₂.

Zarówno wyniki badań, jak i perspektywy dalszego rozwoju zaproponowanych rozwiązań potwierdzają istotny wkład pracy w rozwój dyscypliny naukowej „Inżynieria lądowa, geodezja i transport”.

LITERATURA

- [1] Abediasl, H., Ansari, A., Hosseini, V., Koch, C. R., Shahbakhti, M., 2023. Real-time vehicular fuel consumption estimation using machine learning and on-board diagnostics data. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*. 2023;0(0). doi:10.1177/09544070231185609
- [2] Abukhalil, T., AlMahafzah, H., Alksasbeh, M., Alqaralleh, B. A. 2020. Fuel consumption using OBD-II and support vector machine model. *Journal of Robotics*, 2020, 1-9.
- [3] ACEA Driving mobility for Europe. 2022. New EU car sales by power source. <https://www.acea.auto/figure/fuel-types-of-new-passenger-cars-in-eu>.
- [4] ACEA Driving mobility for Europe. 2023. New EU car sales by power source. <https://www.acea.auto/figure/fuel-types-of-new-passenger-cars-in-eu/>,
- [5] Ahmed, N. S., Sadiq, M. H., 2018. Clarify of the random forest algorithm in an educational field. 2018 International Conference on Advanced Science and Engineering (ICOASE), Duhok, Iraq. 179–184. doi: 10.1109/ICOASE.2018.8548804.
- [6] Alessandrini, A., Cattivera, A., Filippi, F., Ortenzi, F., 2012. Driving style influence on car CO₂ emissions. In *2012 international emission inventory conference*. 1-11.
- [7] Al-Hussein, W. A., Kiah, M. L. M., Yee, L., Zaidan, B. B., 2021. A systematic review on sensor-based driver behaviour studies: coherent taxonomy, motivations, challenges, recommendations, substantial analysis and future directions. *PeerJ Computer Science*, 7, e632.
- [8] Andrych-Zalewska, M., Chlopek, Z., Pielecha, J., Merkisz, J., 2023. Investigation of exhaust emissions from the gasoline engine of a light duty vehicle in the Real Driving Emissions test. *Eksploatacja i Niezawodność – Maintenance and Reliability*, 25(2). doi: 10.17531/ein/165880
- [9] Augustynowicz, A., 2009. *Modelowanie typu kierowcy samochodu*. Oficyna Wydawnicza Politechniki Opolskiej.
- [10] Bai, Z. W., Hao, P., Shangguan, W., Cai, B. G., Barth, M. J., 2022. Hybrid reinforcement learning-based eco-driving strategy for connected and automated vehicles at signalized intersections. *IEEE Transactions on Intelligent Transportation Systems*, 23(9), 15850-15863, doi: 10.1109/TITS.2022.3145798.
- [11] Bakibillah, A. S. M., Kamal, M. A. S., Tan, C. P., Hayakawa, T., Imura, J. I., 2019. Event-Driven Stochastic Eco-Driving Strategy at Signalized Intersections From Self-Driving

- Data. *IEEE Transactions on Vehicular Technology*, 68(9), 8557-8569, doi: 10.1109/TVT.2019.2931519.
- [12] Bifulco, G. N., Galante, F., Pariota, L., Russo Spena, M., 2015. A linear model for the estimation of fuel consumption and the impact evaluation of advanced driving assistance systems. *Sustainability*, 7(10), 14326-14343.
- [13] Borucka, A., 2025. Analiza i modelowanie procesów transportowych, Wydawnictwo Wojskowej Akademii Technicznej, Warszawa.
- [14] Brzeziński, M., 2010. Modelowanie systemu remontu techniki wojsk lądowych. Wojskowa Akademia Techniczna.
- [15] Chan, T. K., Chin, C. S., Chen, H., Zhong, X., 2019. A comprehensive review of driver behavior analysis utilizing smartphones. *IEEE Transactions on Intelligent Transportation Systems*, 21(10), 4444-4475.
- [16] Chen, C., Zhao, X., Yao, Y., Zhang, Y., Rong, J., Liu, X., 2018. Driver's eco-driving behaviour evaluation modeling based on driving events. *Journal of Advanced Transportation*. Doi: 10.1155/2018/9530470.
- [17] Chen, M. C., Yeh, C. T., Wang, Y. S., 2020. Eco-driving for urban bus with big data analytics. *Journal of Ambient Intelligence and Humanized Computing*, doi: 10.1007/s12652-020-02287-2.
- [18] Chen, S.-H., Pan, J.-S., Lu, K., 2015. Driving behavior analysis based on vehicle OBD information and adaboost algorithms. *Proceedings of the International Multiconference of Engineers and Computer Scientists*. 1:18-20.
- [19] Chłopek, Z. 2004. A simulation researches of a pollution emission in a vehicle traffic. In 3rd International Conference Maintenance.
- [20] Chłopek, Z., 2002. Ochrona środowiska naturalnego. Wydawnictwo Komunikacji i Łączności, Warszawa.
- [21] Chłopek, Z., Zimakowska, M., 2010. Ocena sposobów wyznaczania charakterystyk emisji zanieczyszczeń z pojazdów samochodowych w modelach INFRAS i COPERT. *Prace Naukowe Politechniki Warszawskiej serii Inżynieria Środowiska*, Warszawa, 58, 47–60.
- [22] Czasopismo E-magazyny. 2023. Te kraje emitują najwięcej CO2 na świecie, <https://e-magazyny.pl/aktualnosci/oze/te-kraje-emituja-najwiecej-co2-na-swiecie/>.
- [23] Czasopismo Notes from Poland. 2023. <https://notesfrompoland.com/2023/04/06/poland-overtakes-us-to-have-worlds-second-largest-lithium-ion-battery-production-capacity/>

- [24] Czasopismo Warsztat.pl. Diagnostyka według standardu EOBD/ODB II, <https://warsztat.pl/artykuly/diagnostyka-wg-standardu-eobd-obd-ii,61543>.
- [25] Czasopismo Warsztat.pl. Moment obrotowy i moc silnika a jego obciążenie (1), <https://warsztat.pl/dzial/7-czesci-i-regeneracja/artykuly/moment-obrotowy-i-moc-silnika-a-jego-obciazenie-1,56560>.
- [26] Deris, A. M., Zain, A. M., Sallehuddin, R., 2011. Overview of support vector machine in modelling machining performances. *Procedia Eng* 24:308–312.
- [27] Dittmann, P., 2008. Prognozowanie w przedsiębiorstwie. Oficyna Wolters Kluwer bussines, Kraków.
- [28] Dong, Y., Xu, J., Ni, J., 2023. Carbon emission model of vehicles driving at fluctuating speed on highway. *Environmental Science and Pollution Research*, 30(7), 18064-18077.
- [29] Dzhelekarski, P., Alexiev, D., 2005. Initializing communication to vehicle OBDII system. *Electronics*, 5, 21.
- [30] Elektroniczny podręcznik statystyki Statsoft. Dopasowanie rozkładu, https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstdisfit.html.
- [31] Elektroniczny podręcznik statystyki Statsoft. Drzewa klasyfikacyjne, https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstclatre.html.
- [32] Elektroniczny podręcznik statystyki Statsoft. Wzmacniane drzewa klasyfikacyjne i regresyjne, https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstbootres.html.
- [33] Encyclopedia PWN, <https://encyklopedia.pwn.pl/>.
- [34] Ericsson, E., 2001. Independent driving pattern factors and their influence on fuel-use and exhaust emission factors. *Transp. Res. Part D Transp. Environ.* 6, 325–345.
- [35] European Environment Agency. 2024. New registration of electric vehicles in Europe, <https://www.eea.europa.eu/en/analysis/indicators/new-registrations-of-electric-vehicles>.
- [36] Fafoutellis, P., Mantouka, E. G., Vlahogianni, E. I., 2021. Eco-Driving and its impacts on fuel efficiency: An overview of technologies and data-driven methods. *Sustainability*, 13(1):226, doi: 10.3390/su13010226.
- [37] Fontaras, G., Zacharof, N. G., Ciuffo, B., 2017. Fuel consumption and CO₂ emissions from passenger cars in Europe—Laboratory versus real-world emissions. *Progress in energy and combustion Science*, 60, 97-131.

- [38] Gabrysiewicz, M., Zając, P., 2019. Budowa pojazdów samochodowych. Wydawnictwa Komunikacji i Łączności.
- [39] Gatak, P. 2019. Uczenie maszynowe: Klasyfikatory, drzewo decyzyjne, <http://gatak.pl/2019/09/05/klasyfikatory-drzewo-decyzyjne/>.
- [40] Ghosh, D. K., 2024. Perspective Chapter: Linear Regression and Logistic Regression Models. In Recent Advances in Biostatistics. IntechOpen, doi: 10.5772/intechopen.1003183.
- [41] Gronowicz, J., 2003. Ochrona środowiska w transporcie lądowym. Biblioteka problemów eksploatacji. Wydawnictwo i Zakład Poligrafii Instytutu Technologii Eksploatacji. Poznań – Radom.
- [42] Hao, L., Ren, Y., Lu, W., Jiang, N., Ge, Y., Wang, Y., 2023. Assessment of Heavy-Duty Diesel Vehicle NO_x and CO₂ Emissions Based on OBD Data. *Atmosphere*, 14(9), 1417, doi: 10.3390/atmos14091417.
- [43] Hermawan, G.; Husni, E., 2020. Acquisition, modeling, and evaluating method of driving behavior based on OBD-II: A literature survey. 879, 12030.
- [44] Hien, N. L. H., Kor, A. L. 2022. Analysis and prediction model of fuel consumption and carbon dioxide emissions of light-duty vehicles. *Applied Sciences*, 12(2), 803, doi: 10.3390/app12020803.
- [45] <https://www.shutterstock.com/pl/>
- [46] International Monetary Fund. 2024. Inflacja, ceny konsumpcyjne - Unia Europejska, <https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG?locations=EU>.
- [47] Ivanisevic, A.; Radisic, M.; Njegovan, M.; Pavlovic, A., 2020. Development of an Effective Planning Model for Improving Financial Performance. *Forum Sci. Oeconomia* 2020, 8, 67–81, doi:10.23762/FSO_VOL8_NO1_5.
- [48] Izdebski, M., 2018. Modelowanie i analiza problemów decyzyjnych przydziału pojazdów do zadań w zagadnieniach transportowych. Oficyna Wydawnicza Politechniki Warszawskiej.
- [49] Jacyna, M., 2009. Modelowanie i ocena systemów transportowych. Oficyna Wydawnicza Politechniki Warszawskiej.
- [50] Jin, L., Guo, B., Jiang, Y., Hua, Q., 2021. Analysis on the influencing factors of driving behaviours based on the theory of planned behaviour. *Advances in Civil Engineering*, 2021, 6687674, doi: 10.1155/2021/6687674.
- [51] Juda-Rezler, K., 2006. Oddziaływanie zanieczyszczeń powietrza na środowisko. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa.

- [52] Kamińska, J. A., 2018. Residuals in the modelling of pollution concentration depending on meteorological conditions and traffic flow, employing decision trees. In ITM Web of Conferences, 23, 00016, doi: 10.1051/itmconf/20182300016.
- [53] Kamińska, J. A., Chalfen, M., Szczucka-Lasota, B., 2017. Wpływ ruchu samochodowego oraz warunków meteorologicznych na stężenia tlenków azotu. *Autobusy: technika, eksploatacja, systemy transportowe*, 18(12), 196-200.
- [54] Karczewski, M. E., Szczęch, L. P., Trawiński, G., 2013. Silniki pojazdów samochodowych. Podręcznik do nauki zawodu technik pojazdów samochodowych. Wydawnictwa Szkolne i Pedagogiczne.
- [55] Kashevnik, A., Lashkov, I., Gurtov, A., 2020. Methodology and mobile application for driver behavior analysis and accident prevention. *IEEE Transactions on Intelligent Transportation Systems*, 21(6), 2427-2436, doi: 10.1109/TITS.2019.2918328.
- [56] Khan, A. A., Ali, M. H., Haque, F., Habib, T., 2023. A Machine Learning Approach for Driver Identification. *Indonesian Journal of Electrical Engineering and Computer Science*, 30(1), April 2023, 276~288 ISSN: 2502-4752, doi: 10.11591/ijeecs.v30.i1.pp276-288.
- [57] Khedkar, S., Oswal, A., Setty, M., Ravi, S., 2015. Driver evaluation system using mobile phone and OBD-II system. *International Journal of Computer Science and Information Technologies*, 6(3), 2738-2745.
- [58] Kilian, L., Zhou, X., 2022. The Impact of Rising Oil Prices on U.S. Inflation and Inflation Expectations in 2020–23. *Energy Economics*, 2022, 113, 106228, doi: 10.1016/j.eneco.2022.106228.
- [59] Komisja Europejska. 2022. Baza danych dotycząca energii i środowiska, <https://transport.ec.europa.eu/system/files/2022-09/pb2022-part3.xlsx>.
- [60] Komisja Unii Europejskiej. 2023. Emisja gazów GHG na świecie, https://edgar.jrc.ec.europa.eu/report_2023.
- [61] Kryszicki, W., Bartos, J., Dyczka, W., Królikowska, K., Wasilewski, M., 1995) *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach, część II. Statystyka matematyczna*, Wydawnictwo Naukowe PWN, Warszawa.
- [62] Kułakowski P. 2024. Regresja logistyczna – materiały dydaktyczne, https://home.agh.edu.pl/~mmd/_media/dydaktyka/adp/regresja_logistyczna.pdf
- [63] Kułakowski, P. 2024. Regresja logistyczna, problemy bias/variance, diagnostyka, http://home.agh.edu.pl/~mmd/_media/dydaktyka/adp/regresja_logistyczna.pdf.

- [64] Kumar, R., Jain, A., 2022. Monitoring and remote data logging of engine operation via on board diagnostic port. 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT), Sonapat, India, 2022, 550-555, doi: 10.1109/CCICT56684.2022.00102.
- [65] Kumar, R., Jain, A., 2023. Driving behavior analysis and classification by vehicle OBD data using machine learning. *The Journal of Supercomputing*, 79, 18800–18819, doi: 10.1007/s11227-023-05364-3.
- [66] Kwak, B. I., Woo, J., Kim, H. K., 2016. Know your master: driver profiling-based anti-theft method. 2016 14th Annual Conference on Privacy, Security and Trust (PST), Auckland, New Zealand, 2016, 211-218, IEEE, doi: 10.1109/PST.2016.7906929.
- [67] Lasocki, J., Chłopek, Z., Godlewski, T., 2019. Driving style analysis based on information from the vehicle's OBD system. *Combustion Engines*, 178(3), 173-181, doi: 10.19206/CE-2019-330 58.
- [68] Lotko, W., 2005. Wybrane zagadnienia diagnostyki pojazdów. Politechnika Radomska, Radom.
- [69] Ma, Z., Jørgensen, B. N., Ma, Z., 2024. A Scoping Review of Energy-Efficient Driving Behaviors and Applied State-of-the-Art AI Methods. *Energies*, 17(2), 500, doi: 10.3390/en17020500.
- [70] Maciąg, A., Kukła, S., Pietroń, R., 2013. Prognozowanie i symulacja w przedsiębiorstwie. Polskie Wydawnictwo Ekonomiczne, Warszawa.
- [71] Mamczur, M. 2024. Blog o data science, AI, uczeniu maszynowym i wizualizacji danych. Czym jest i jak działa drzewo decyzyjne?, <https://mirosławmamczur.pl/czym-jest-i-jak-działa-drzewo-decyzyjne/>
- [72] Martinelli, F., Marulli, F., Mercaldo, F., Santone, A., 2021. Neural networks for driver behavior analysis. *Electronics*, 10(3), 342, doi:10.3390/electronics10030342.
- [73] Massoud, R., Bellotti, F., Berta, R., De Gloria, A., Poslad, S., 2019. Eco-driving profiling and behavioral shifts using IoT vehicular sensors combined with serious games. 2019 IEEE Conference on Games (CoG), London, UK, 2019, 1-8, doi: 10.1109/CIG.2019.8847992.
- [74] Mądziel, M., 2023. Vehicle Emission Models and Traffic Simulators: A Review. *Energies*, 16(9), 3941, doi: 10.3390/en16093941.
- [75] Mądziel, M., Jaworski, A., Kuszewski, H., Woś, P., Campisi, T., Lew, K., 2021. The development of CO₂ instantaneous emission model of full hybrid vehicle with the use of machine learning techniques. *Energies*, 15(1), 142, doi: 10.3390/en15010142.

- [76] Merkisz, J., 1999. Ekologiczne problemy silników spalinowych Tom I i Tom II. Wydawnictwo Politechniki Poznańskiej, Poznań.
- [77] Merkisz, J., Dobrzynski, M., Kubiak, K., 2017. An impact assessment of functional systems in vehicles on CO₂ emissions and fuel consumption. *MATEC Web of Conferences* VII International Congress on Combustion Engines, 118. EDP Sciences, doi: 10.1051/mateconf/201711800030.
- [78] Merkisz, J., Mazurek, S., 2007. Pokładowe systemy diagnostyczne pojazdów samochodowych. Wydawnictwa Komunikacji i Łączności, Warszawa.
- [79] Meseguer, J. E., Calafate, C. T., Cano, J. C., Manzoni, P., 2013. Driving Styles: A Smartphone Application to Assess Driver Behavior. 2013 IEEE Symposium on Computers and Communications (ISCC), Split, Croatia, 000535-000540, doi: 10.1109/ISCC.2013.6755001.
- [80] Meseguer, J. E., Calafate, C. T., Cano, J. C., Manzoni, P., 2015. Assessing the impact of driving behavior on instantaneous fuel consumption. 2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, USA, 443-448, doi: 10.1109/CCNC.2015.7158016.
- [81] Oduro, S. D., Metia, S., Duc, H., Ha, Q. P., 2013. CO₂ vehicular emission statistical analysis with instantaneous speed and acceleration as predictor variables. 2013 International Conference on Control, Automation and Information Sciences (ICCAIS), Nha Trang, Vietnam, 158-163, doi: 10.1109/ICCAIS.2013.6720547.
- [82] Osowski, S., 2013. *Metody i narzędzia eksploracji danych*. Wydawnictwo BTC.
- [83] Pan, Y. J., Yu, T. C., Cheng, R. S., 2017. Using OBD-II data to explore driving behavior model. 2017 International Conference on Applied System Innovation (ICASI), Sapporo, Japan, 1816-1818, doi: 10.1109/ICASI.2017.7988297.
- [84] Parlament Europejski. 2024. Emisje CO₂ z samochodów: fakty i mity, https://www.europarl.europa.eu/pdfs/news/expert/2019/4/story/20190313STO31218/20190313STO31218_pl.pdf.
- [85] Parlament Europejski. 2024. Infografika: Emisje gazów cieplarnianych w Unii Europejskiej według sektora i kraju, https://www.europarl.europa.eu/pdfs/news/expert/2018/3/story/20180301STO98928/20180301STO98928_pl.pdf.
- [86] Peppes, N., Alexakis, T., Adamopoulou, E., Demestichas, K., 2021. Driving behaviour analysis using machine and deep learning methods for continuous streams of vehicular data. *Sensors*, 21(14), 4704, doi:10.3390/s21144704.

- [87] Ping, P., Qin, W., Xu, Y., Miyajima, C., Takeda, K., 2019. Impact of driver behavior on fuel consumption: Classification, evaluation and prediction using machine learning. In IEEE Access, 7, 78515-78532, doi: 10.1109/ACCESS.2019.2920489.
- [88] Pinto, G., Oliver-Hoyo, M. T., 2008. Using the relationship between vehicle fuel consumption and CO₂ emissions to illustrate chemical principles. Journal of Chemical Education 85 (2), 218, doi: 10.1021/ed085p218.
- [89] PKN Orlen. Baza danych o cenach paliw, <https://www.orsen.pl/en/for-business/fuel-wholesale-prices>.
- [90] Płaszczycza, M. Statystyka od A do Z: Metoda CART, <https://www.statystyka.eu/analiza-skupien/metoda-cart.php>.
- [91] Polski Związek Przemysłu Motoryzacyjnego. 2023. Raport ACEA “Pojazdy w użyciu: Europa”, <https://www.pzpm.org.pl/pl/content/download/15673/108016/file/ACEA-report-vehicles-in-use-europe-2023.pdf>.
- [92] Polski Związek Przemysłu Motoryzacyjnego. 2023. Raport ACEA “Pojazdy w użyciu”, https://www.pzpm.org.pl/pl/Publikacje/NEWSLETTER/Newsletter-STYCZEN-2023/01_Raport-ACEA-Pojazdy-w-uzyciu-z-2023.
- [93] Powierża, L., 1997. Elementy teorii systemów. Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej.
- [94] Prochowski, L., 2016. Mechanika Ruchu, Wydawnictwa Komunikacji i Łączności.
- [95] Puchalski, A., Komorska, I., 2020. Driving style analysis and driver classification using OBD data of a hybrid electric vehicle. Transport Problems, 15(4, Part 1):83-94, doi: 10.21307/tp-2020-050.
- [96] Rabiej, M., 2012. Statystyka z Programem Statistica (Statistics in Statistica Software). Helion, Gliwice, ISBN 9788324641109.
- [97] Rudak, O. 2020. Tolerancja prędkościomierza samochodowego. Czasopismo Lege Artis, <https://czasopismo.legeartis.org/2020/01/tolerancja-predkosciomierza-tachometru-samochodowego-wzor-odstepstwo-rzeczywistej-blad-pomiarowy/>.
- [98] Rykała, M., Grzelak, M., Rykała, Ł., Voicu, D., Stoica, R. M., 2023. Modeling Vehicle Fuel Consumption Using a Low-Cost OBD-II Interface. Energies, 16(21), 7266, doi: 10.3390/en16217266.
- [99] Rymaniak, Ł., Merkiś, J., Szymlet, N., Kamińska, M., Weymann, S., 2021. Use of emission indicators related to CO₂ emissions in the ecological assessment of an agricultural tractor. Eksploatacja i Niezawodność – Maintenance and Reliability, 23(4), 605-611, doi: 10.17531/ein.2021.4.2.

- [100] Sawczuk, W., Merkisz-Guranowska, A., Rilo Cañas, A., 2021,. Assessment of disc brake vibration in rail vehicle operation on the basis of brake stand. *Eksploracja i Niezawodność – Maintenance and Reliability*, 23(2), 221-230, doi: 10.17531/ein.2021.2.2.
- [101] Sawczuk, W., Merkisz-Guranowska, A., Rilo Cañas, A., Kołodziejski, S. 2022. New approach to brake pad wear modelling based on test stand friction-mechanical investigations. *Eksploracja i Niezawodność – Maintenance and Reliability*, 24(3), 3, doi: 10.17531/ein.2022.3.3
- [102] Shaikh, M. K., Palaniappan, S., Ali, E. F., Khurram, E. D. M., 2020. Identifying Driver Behaviour Through OBD-II Using Android Application. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 17(7), 13636-13647.
- [103] Shpak, N., Podolchak, N., Karkovska, V., Sroka, W., Horbal, N., 2022. The Application of Tools for Assessing the Financial Security of Enterprises. *Forum Scientiae Oeconomia*, 10, 29–44, doi: 10.23762/FSO_VOL10_NO2_2.
- [104] Singh, S. K., Singh, A. K., 2022. Driving pattern analysis to determine driver behaviors for local authority based on cloud using OBD II. *International Journal of electrical and computer engineering systems*, 13(10), 937-944, doi 10.32985/ijeces.13.10.14:.
- [105] Sklep Autocentrum. Dane techniczne pojazdu Mazda 3. <https://www.autocentrum.pl/dane-techniczne/mazda/3/iii/sedan/silnik-benzynowy-2.0-skyactiv-g-120km-2013-2016/>,
- [106] Sklep ebay. Zdjęcie Złącza diagnostyczne OBD-II DLC Pigtail GM 96-19 OBD 2 (różne długości), <https://www.ebay.ca/itm/203317536688>.
- [107] Sklep Empik. Zdjęcie uniwersalnego czytnika kodów OBD II, <https://www.empik.com/uniwersalny-czytnik-kodow-obd-ii-technaxx-tx-184,p1370307650,motoryzacja-p>.
- [108] Sklep Interface24. 2024. Zdjęcie iCar Pro Wifi, https://interface24.pl/produkt/elm327-icar-pro-wifi-vgate-obd2-interfejs-ios-android-pl/?gad_source=1&gclid=CjwKCAiAlcyuBhBnEiwAOGZ2S-KZJbTsLihZh8xvWAtwXtBjSpu3YEKQ9wL23dSx-LG4UIvBtdysbBoCtzQQA_vD_BwE
- [109] Sklep Interface24. 2024. Zdjęcie komputera diagnostycznego, <https://interface24.pl/produkt/24diag-s403-komputer-interfejs-diagnostyczny-obd-2-uniwersalny/>.

- [110] Sklep Interface24. 2024. Zdjęcie skaner-tester diagnostyczny, <https://interface24.pl/produkt/launch-creader-elite-ford-mazda-interfejs-skaner-tester-diagnostyczny>.
- [111] Stanisław, A., 2007. Przystępny kurs statystyki: z zastosowaniem STATISTICA PL na przykładach z medycyny. Analizy wielowymiarowe. StatSoft.
- [112] Stanisław, A., 2016. Modele regresji logistycznej. Zastosowanie w medycynie, naukach przyrodniczych i społecznych. StatSoft Polska, Kraków.
- [113] Suzdaleva, E.; Nagy, I. Data-based speed-limit-respecting eco-driving system. *Transportation Research Part C: Emerging Technologies*, 44, 253-264, doi: 10.1016/j.trc.2014.04.009 .
- [114] Syahputra, R., 2016. Application of neuro-fuzzy method for predictions for prediction of vehicle fuel consumption. *Journal of Theoretical and Applied Information Technology*, 86(1), 138–150.
- [115] System elektronicznego poboru opłaty drogowej, <https://mytoz.eu/pl/co2>.
- [116] Szczęsny, P., Rymarz, J., 2012. Wskaźniki bezpieczeństwa ruchu drogowego w Polsce i krajach sąsiednich. *Autobusy: technika, eksploatacja, systemy transportowe*, 13, 171-178.
- [117] Szokało, A. A., Rychter, M., 2018. Rozwój systemu diagnostyki pokładowej samochodów osobowych w świetle zasady działania systemu OBD. *Autobusy: technika, eksploatacja, systemy transportowe*, 19(6), doi: 10.24136/atest.2018.167 .
- [118] Świdorski, A., 2011. Modelowanie oceny jakości usług transportowych. *Prace Naukowe Politechniki Warszawskiej. Transport*, 81, 3-160.
- [119] Tarnowski, W., Bartkiewicz, S., 1998. Modelowanie matematyczne i symulacja komputerowa dynamicznych procesów ciągłych. Wydawnictwo FENIKS, Koszalin.
- [120] Uvarov, K., Ponomarev, A., 2021. Driver identification with OBD-II public data. 2021 28th Conference of Open Innovations Association (FRUCT), Moscow, Russia, 495-501, doi: 10.23919/FRUCT50888.2021.9347648.
- [121] Villavicencio, C. N., Jeng, J.-H., Hsieh, J.-G., 2021. Support vector machine modelling for COVID-19 prediction based on symptoms using R programming language. In *Proceedings of the 2021 4th International Conference on Machine Learning and Machine Intelligence (MLMI '21)*. Association for Computing Machinery, New York, USA, 65–70, doi: 10.1145/3490725.3490735.
- [122] Wasilewska, E., 2011. Statystyka opisowa od podstaw: podręcznik z zadaniami. Wydawnictwo SGGW.

- [123] Wickramanayake, S., Dilum Bandara, H. M. N., 2016. Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka, 90-95, doi: 10.1109/MERCon.2016.7480121.
- [124] Witaszek, K., 2020. Modeling of fuel consumption using artificial neural networks. *Diagnostyka*, 21(4), 103–113, doi: 10.29354/diag/130610.
- [125] World Bank national accounts data. 2024. Produkt Krajowy Brutto – Polska, <https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=PL>.
- [126] World Bank Open Data, <https://data.worldbank.org/>.
- [127] Zhao, D., Li, H., Hou, J., Gong, P., Zhong, Y., He, W., Fu, Z., 2023. A Review of the Data-Driven Prediction Method of Vehicle Fuel Consumption. *Energies*, 16(14), 5258, doi: 10.3390/en16145258
- [128] Zimakowska-Laskowska, M., 2014. Model emisji zanieczyszczeń z silników pojazdów samochodowych do atmosfery. Praca doktorska. Politechnika Warszawska.
- [129] Ziółkowski, J., Oszczypała, M., Małachowski, J., Szkutnik-Rogoż, J., 2021. Use of artificial neural networks to predict fuel consumption on the basis of technical parameters of vehicles. *Energies*, 14(9), 2639, doi: 10.3390/en14092639.
- [130] ZS Koszalin. Zdjęcie interfejsu systemu OBD2 i EOBD, <https://www.samochodowka.koszalin.pl/warsztaty/dzialy/bielicki/obd/systemobd2-eobd.htm>.