



**POLITECHNIKA WARSZAWSKA**  
**WYDZIAŁ**  
**MECHANICZNY ENERGETYKI I LOTNICTWA**  
**ZAKŁAD SILNIKÓW LOTNICZYCH**

---



**PRACA PRZEJŚCIOWA**

PRACA INŻYNIERSKA

Filip Rusiecki

Zastosowanie modeli uczenia maszynowego do wstępnej fazy projektu silnika spalinowego

299153

Lotnictwo i Kosmonautyka  
Napędy lotnicze

**Promotor:** dr inż. Mateusz Żbikowski

Warszawa, wrzesień 2021

## **Streszczenie:**

Niniejsza praca, podejmuje problematykę dotyczącą wstępnej fazy projektu silnika spalinowego. Praca opiera się na podstawowych modelach uczenia maszynowego do predykcji takich cech silnika jak zużycie paliwa [km/l] oraz tego czy silnik potrzebuje turbodoładowania, sprężarki doładowującej lub systemu Start Stop. Niniejsza praca, przedstawia także rozważania na przyszłe możliwe zastosowania uczenia maszynowego w inżynierii.

## Spis treści:

Wstęp.....	3
1. Cel i założenie projektu.....	4
1.1. Modele uczenia maszynowego.....	4
1.1.1. Regresja logistyczna.....	4
1.1.2. Regresja za pomocą Lasu Losowego .....	5
1.2. Przewidywane własności silnika .....	6
1.2.1. Zużycie paliwa [km/l] .....	6
1.2.2. Turbodoładowanie.....	6
1.2.3. Sprężarka doładowująca.....	7
1.2.4. System Start Stop .....	7
2. Implementacja .....	7
2.1. Schemat działania.....	7
2.2. Regresja logistyczna.....	8
2.3. Regresja za pomocą Lasu Losowego .....	8
3. Wyniki.....	9
3.1. Zużycie paliwa [km/l] .....	9
3.2. Turbodoładowanie.....	9
3.3. Sprężarka doładowująca.....	10
3.4. System Start Stop .....	10
Podsumowanie .....	11
Literatura .....	12

## Wstęp

Rozwój techniki wraz z postępującym rozwojem kart graficznych spowodował wzrost zainteresowania zastosowaniami modeli uczenia maszynowego. Modele te znacząco uproszczają złożoność rozwiązywanego zagadnienia, przez co zmniejszają czas oraz koszty, pozwalają też szukać nowych nieodkrytych dotychczas rozwiązań.

Uczenie maszynowe jest jedną z najprężniej rozwijających się nauk w dzisiejszych czasach, wielkie korporacje, zauważyły gigantyczne ekonomiczne możliwości płynące z niego. Używa się go praktycznie wszędzie tam, gdzie potrzeba jest wykrycia wzorców lub podjęcia decyzji, które nie są oczywiste. Na przykład w medycynie używa się jej do wykrycia raka skóry lub zakwalifikowania złamanych kości. Google używa głębokiego uczenia do personalizacji reklam. Ta gałąź zdecydowanie rośnie w siłę, wobec czego pojawia się co raz więcej chętnych do coraz szerszych zastosowań. Ponadto większy popyt na gromadzenie dużej ilości danych (Big Data) pozwolił również rozwinąć uczenie maszynowe na działy trudniejsze do implementacji, w tym inżynierię. Dlatego też, w niniejszej pracy podjęto się zastosowania podstawowych modeli uczenia maszynowego do predykcji cech silnika, za pomocą kilku jego podstawowych własności. Dzięki temu można w prosty sposób, przyspieszyć wstępną fazę projektu silnika. W pierwszym rozdziale zawarto opis użytych modeli. Rozdział drugi poświęcony jest implementacji modeli. W ostatnim rozdziale przedstawiono wyniki predykcji danych cech silników. W podsumowaniu dokonano również analizy rynku pod względem istniejących już zastosowań w inżynierii.

# 1. Cel i założenie projektu

Na cel niniejszej pracy składa się zastosowanie dwóch modeli uczenia maszynowego do predykcji podstawowych cech i własności silnika spalinowego. Pierwszym z nich jest regresja za pomocą Lasu losowego, pozwala nam ona na predykcję zużycie paliwa [km/l]. Drugim natomiast jest Regresja logistyczna, dzięki której można przewidzieć czy silnik potrzebuje turbodoładowania, sprężarki doładowującej lub systemu Start Stop. Do napisania kodu został użyty język programowania Python, a w nim biblioteki do uczenia maszynowego takie jak Keras czy Sci-kit learn. Założeniem projektu jest ukazanie dużego zastosowania uczenia maszynowego w inżynierii, ponieważ pomimo wielu zalet, firmy wciąż boją się wykorzystywać uczenie maszynowe z powodu zbyt złożonych i niezrozumiałych wewnętrznych procesów.

## 1.1. Modele uczenia maszynowego

### 1.1.1. Regresja logistyczna

Regresja logistyczna jest jedną z najczęściej stosowanych w badaniach procedur statystycznych. Jest składnikiem prawie wszystkich komercyjnych pakietów statystycznych i jest uważana za jedną z najważniejszych modeli. W przeciwieństwie do tradycyjnej regresji liniowej, regresja logistyczna jest odpowiednia do modelowania zmiennej binarnej, co idealnie wpasowuje się w wstępną fazę projektu silnika, gdzie trzeba zdecydować czy używać konkretnej technologii. Regresja logistyczna jest używana przede wszystkim do modelowania zmiennej binarnej w oparciu o jedną lub więcej innych zmiennych, zwanych predyktorami. Modelowana zmienna binarna jest nazywana zmienną zależną. Model regresji logistycznej oblicza ważoną sumę cech wejściowych. wyprowadza logistykę tego wyniku. Prawdopodobieństwo modelu regresji logistycznej (postać wektoryzowana ukazana w równaniu 1.1.1.1)

$$\hat{p} = h_{\theta}(\mathbf{x}) = \sigma(\mathbf{x}^T \boldsymbol{\theta}) \quad [\text{równanie 1.1.1.1}]$$

Prawdopodobieństwo to oznaczone  $\sigma(\cdot)$  jest funkcją sigmoidalną (równanie 1.1.1.2), której wynikiem jest liczba od 0 do 1.

$$\sigma(t) = \frac{1}{1 + e^{-t}} \quad [\text{równanie 1.1.1.2}]$$

Gdy model regresji logistycznej oszacował prawdopodobieństwo (równanie 1.1.1.1), takie że  $\mathbf{x}$  należy do klasy dodatniej, wtedy może dokonać łatwo predykcji  $\hat{y}$  (równanie 1.1.1.3).

$$\hat{y} = \begin{cases} 0 & \text{jeżeli } \hat{p} < 0.5 \\ 1 & \text{jeżeli } \hat{p} \geq 0.5 \end{cases} \quad [\text{równanie 1.1.1.3}]$$

Trening modelu opiera się na ustawieniu wektora parametrów  $\boldsymbol{\theta}$  tak, aby model oszacował wysokie prawdopodobieństwa dla przypadków pozytywnych ( $y = 1$ ) i niskie prawdopodobieństwa dla przypadków negatywnych ( $y = 0$ ). Ten zamysł jest uchwycony przez funkcję kosztu pokazaną w równaniu 1.1.1.4 dla przykładowego  $\mathbf{x}$ .

$$c(\boldsymbol{\theta}) = \begin{cases} -\log(\hat{p}) & \text{jeżeli } y = 1 \\ -\log(1 - \hat{p}) & \text{jeżeli } y = 0 \end{cases} \quad [\text{równanie 1.1.1.4}]$$

Funkcja kosztu w całym uczeniu to średni koszt wszystkich przykładów danych, na których model się uczy. Można go zapisać w jednym wyrażeniu zwanym stratą logarytmiczną, pokazanym w równaniu 1.1.1.5

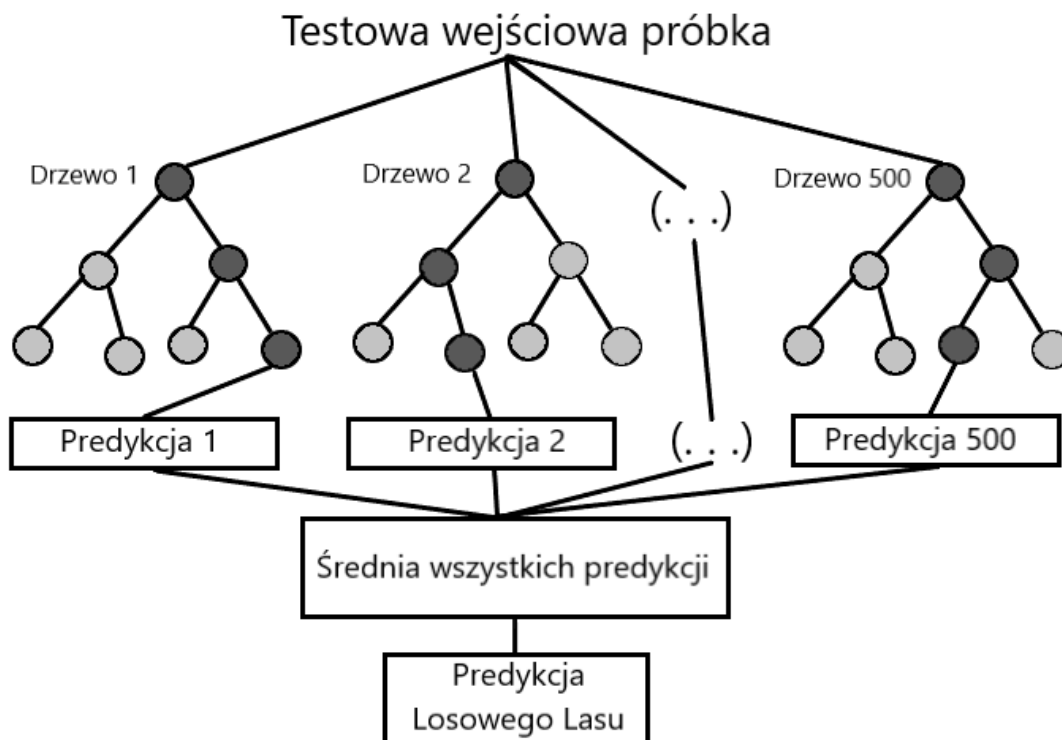
$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log(\hat{p}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{p}^{(i)})] \quad [\text{równanie 1.1.1.5}]$$

Ta funkcja kosztu jest wypukła, więc metoda gradientu prostego gwarantuje znalezienie globalnego minimum. Pochodne cząstkowe funkcji kosztu względem parametru modelu  $\theta_j$  dane są równaniem 1.1.1.6.

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{1}{m} \sum_{i=1}^m (\sigma(\theta^T \mathbf{x}^{(i)}) - y^{(i)}) x_j^{(i)} \quad [\text{równanie 1.1.1.6}]$$

## 1.1.2. Regresja za pomocą Lasu Losowego

Losowy las to meta-estymator, który dopasowuje szereg klasyfikujących drzew decyzyjnych do różnych podpróbek zbioru danych i wykorzystuje uśrednianie w celu poprawy dokładności predykcyjnej i kontroli nadmiernego dopasowania. Rozmiar podpróbki jest kontrolowany za pomocą parametru maksymalnych próbek. Ponadto Losowy Las jest również bardzo szybki i niezawodny w porównaniu do innych modeli regresji. Algorytm losowego lasu łączy dane wyjściowe wielu drzew decyzyjnych w celu wygenerowania ostatecznego wyniku. Regresja losowego lasu to nadzorowany algorytm uczenia, który wykorzystuje metodę uczenia zespołowego do regresji. Metoda uczenia zespołowego to technika, która łączy prognozy z wielu algorytmów uczenia maszynowego, aby uzyskać dokładniejsze prognozy niż pojedynczy model.



1.1.2.1. Diagram przykładowej struktury Losowego Lasu

Powyższy diagram przedstawia strukturę Losowego Lasu. Drzewa biegną równolegle bez interakcji między sobą. Losowy las działa poprzez konstruowanie kilku drzew decyzyjnych w czasie szkolenia i wyprowadzając średnią klas jako predykcję wszystkich drzew. Kolejne kroki działania Losowego Lasu:

1. Wybiera losowo  $k$ -punktów danych ze zbioru treningowego.
2. Buduje drzewo decyzyjne powiązane z tymi  $k$ -punktami danych.
3. Wybiera liczbę  $N$ -drzew, które zbuduje i powtarza kroki 1 i 2.
4. W przypadku nowego punktu danych sprawia, aby każde z drzew  $N$ -drzewa przewidywało wartość  $y$  dla danego punktu danych i przypisuje nowy punkt danych do średniej ze wszystkich przewidywanych wartości  $y$ .

Wady są następujące: brak możliwości interpretacji, łatwo może wystąpić overfitting, musimy wybrać liczbę drzew, które mają być uwzględnione w modelu.

## **1.2. Przewidywane własności silnika**

### **1.2.1. Zużycie paliwa [km/l]**

Zużycie paliwa przez pojazdy jest istotnym czynnikiem zanieczyszczenia powietrza, a import paliwa silnikowego może stanowić dużą część handlu zagranicznego kraju, z tego też powodu wiele krajów nakłada wymagania dotyczące oszczędności paliwa. W celu przybliżenia rzeczywistych osiągnięć pojazdu stosuje się różne metody. Energia zawarta w paliwie jest potrzebna do pokonania różnych strat napotykanych podczas napędzania pojazdu oraz do zasilania układów pojazdu, takich jak zapłon czy klimatyzacja. Z tego powodu obliczenie analityczne zużycia paliwa jest wysoko skomplikowane z powodu nieliniowości problemu. Zatem modele uczenia maszynowego pozwalają nam w wysokiej precyzji oszacować zużycie paliwa bez konieczności posiadania prototypu silnika.

### **1.2.2. Turbodoładowanie**

Turbosprężarki należą do kategorii turbin gazowych o małych rozmiarach. Stosowane są w samochodach osobowych i ciężarówkach. Ich zadaniem jest zwiększenie efektywnego średniego ciśnienia tłoka silników spalinowych, a tym samym podniesienie ich sprawności cieplnej. Spaliny z silnika trafiają do turbiny, gdzie cała ich energia jest częściowo zamieniana na moc. Turbina napędza stopień sprężarki, który zwykle jest jednostopniową sprężarką promieniową. Sprężarka zasysa powietrze z otoczenia, spręża je i dostarcza do tłoka, tym samym znacznie zwiększając średnie efektywne ciśnienie tłoka i sprawność cieplną silnika.

### 1.2.3. Sprężarka doładowująca

Sprężarka doładowująca to sprężarka, która zwiększa ciśnienie lub gęstość powietrza dostarczanego do silnika spalinowego. Daje to w każdym cyklu dolotowym do silnika więcej tlenu, co pozwala spalać więcej paliwa i wykonywać więcej pracy, zwiększając w ten sposób moc wyjściową. Zasilanie doładowania może być dostarczane mechanicznie za pomocą pasa, wału lub łańcucha połączonego z wałem korbowym silnika. Wadą doładowania jest to, że sprężanie powietrza zwiększa jego temperaturę.

Gdy w silniku spalinowym stosuje się sprężarkę doładowującą, temperatura mieszanki paliwowo-powietrznej staje się głównym czynnikiem ograniczającym osiągi silnika. Ekstremalne temperatury spowodują detonację mieszanki paliwowo-powietrznej (silniki o zapłonie iskrowym) i uszkodzenie silnika. Utrzymywanie chłodnego powietrza, które dostaje się do silnika, jest ważną częścią konstrukcji. Sprężanie powietrza zwiększa jego temperaturę, dlatego często stosuje się mały radiator zwany intercoolerem między pompą a silnikiem, aby obniżyć temperaturę powietrza.

### 1.2.4. System Start Stop

Samochodowy system start-stop automatycznie wyłącza i ponownie uruchamia silnik spalinowy, aby skrócić czas pracy silnika na biegu jałowym, zmniejszając w ten sposób zużycie paliwa i emisję spalin. Jest to najkorzystniejsze dla pojazdów, które spędzają dużo czasu czekając na światłach lub często zatrzymują się w korkach. Technologia Start-Stop może stać się bardziej powszechna wraz z bardziej rygorystycznymi rządowymi przepisami dotyczącymi oszczędności paliwa i emisji. W przypadku pojazdów nieelektrycznych korzyści wynikające z zastosowania tej technologii wynoszą zwykle od 3 do 10 procent, a potencjalnie nawet do 12 procent. Silnik nie zatrzyma się, jeśli samochód jest w ruchu.

## 2. Implementacja

### 2.1. Schemat działania

Schemat działania implementacji modeli uczenia maszynowego można opisać w następujący sposób:

1. Sformułowanie problemu – ustalenie co chcemy przewidzieć, jakie dane to dane niezależne, a które to dane zależne.
2. Zebranie danych – ustalenie, które bazy danych będą odpowiednie do naszego problemu.
3. Wstępne przetwarzanie danych – czyszczenie danych, agregacja danych, łączenie baz danych, sprawdzenie korelacji danych, normalizacja danych etc.
4. Podział danych na część treningową, testową i walidacyjną. Podział danych na X i Y.
5. Wybór modelu w zależności od zadanego problemu.
6. Ewaluacja modelu – sprawdzenie modelu za pomocą metryk statystycznych na danych podzielonych na część testową.
7. Dostosowanie parametrów modelu – dostrojenie wag modelu, zmiana bazy danych (np. SMOTE), etc.
8. Powrót do 5,6. Jeżeli wartości są satysfakcjonujące zakończyć pętlę.
9. Sprawdzenie modelu na części walidacyjnej (W wypadku niezadawalających wyników wrócić do 5.).

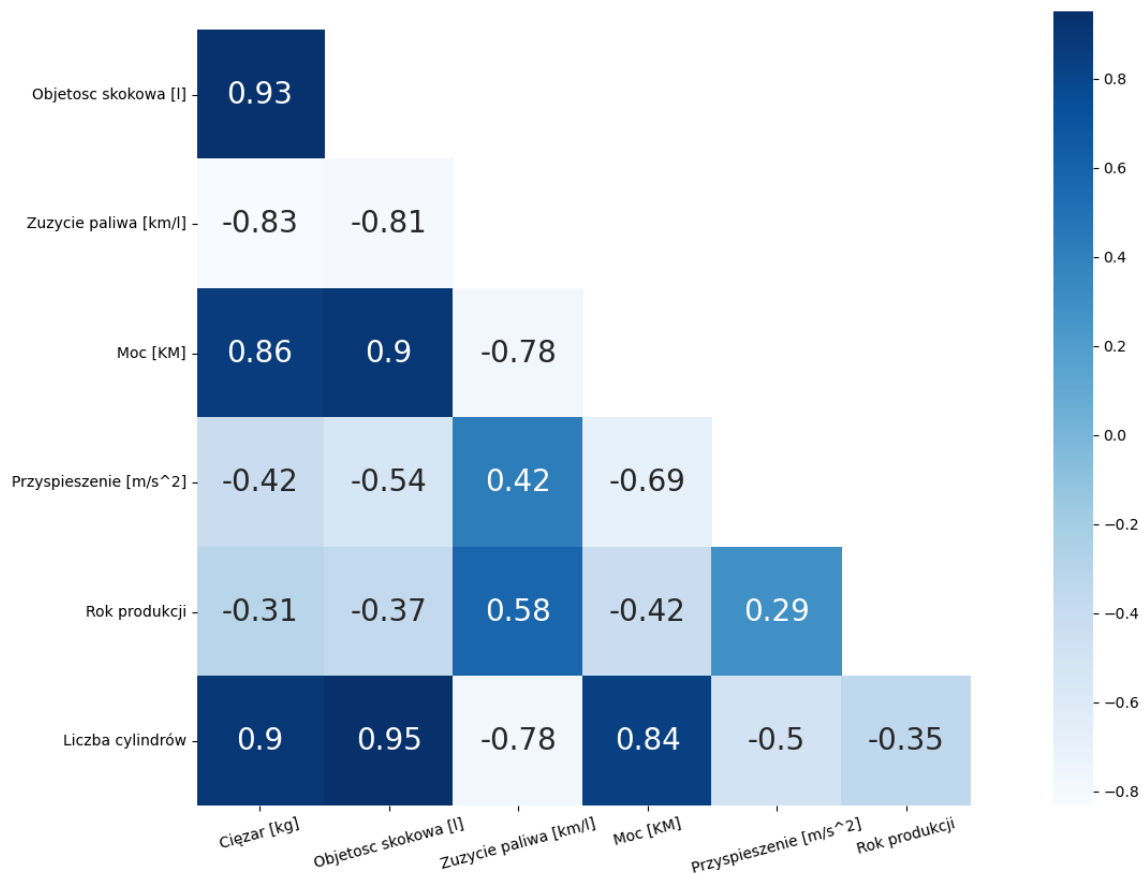


## 2.2. Regresja logistyczna

Regresją logistyczną chcemy określić czy nasz silnik potrzebuje turbodoładowania, sprężarki doładowującej lub systemu Start Stop. Użyte dane zostały zebrane z rządowej strony fueleconomy.gov [5]. Zostały one podzielone na część treningową (70%), testową (20%) oraz walidacyjną (10%). Właściwości samochodu, które zostały użyte do predykcji to rok produkcji, typ samochodu, napęd samochodu, liczba cylindrów, objętość skokowa, typ używanego paliwa oraz zużycie paliwa. Dane zostały oczyszczone z wartości typu null oraz zamieniono wartości typu string na wartości binarne, by przygotować je dla modelu. Podzielono dane na X i Y. Sprawdzono model za pomocą dokładności, precyzji oraz czułości. Do poprawy wyników predykcji turbodoładowania użyto metody SMOTE, która pozwala na lepsze wyniki w przypadku, gdy dane są niezbalansowane. Do pozostałych predykcji użyto zmiany wag, proporcjonalnie do stosunku danych.

## 2.3. Regresja za pomocą Lasu Losowego

Regresją za pomocą Lasu Losowego chcemy określić ile nasz silnik zużywa paliwa [km/l]. Użyte dane zostały zebrane z archiwum uczelni UCL [6]. Zostały one podzielone na część treningową (70%), testową (20%) oraz walidacyjną (10%). Właściwości samochodu, które zostały użyte do predykcji to rok produkcji, liczba cylindrów, objętość skokowa, moc, przyspieszenie samochodu oraz jego ciężar. Dane zostały oczyszczone z wartości typu null oraz zamieniono wartości typu string na wartości binarne, by przygotować je dla modelu. Sprawdzono korelacje danych, przedstawiono je poniżej w macierzy korelacji (2.3.1.). Podzielono dane na X i Y. Sprawdzono model za pomocą dokładności, błędu średniokwadratowego oraz odchylenia standardowego błędu średniokwadratowego.



2.3.1. Macierz korelacji predykcji zużycia paliwa.

### 3. Wyniki

#### 3.1. Zużycie paliwa [km/l]

Dokładność predykcji zużycia paliwa za pomocą Losowego lasu wyniosła 97,97% na próbce treningowej, 85,24% na próbce testowej i 85,76% na próbce walidacyjnej.

Błąd średniokwadratowy predykcji zużycia paliwa dla próbki treningowej wyniósł 0,322, dla próbki testowej 0,879, dla walidacyjnej 0,879.

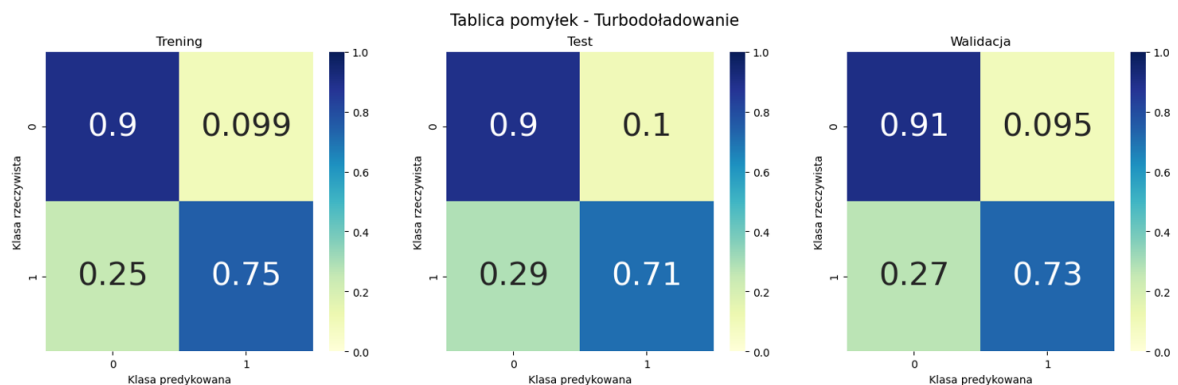
Odchylenie standardowe błędu średniokwadratowego predykcji zużycia paliwa dla próbki treningowej wyniosło 0,483, dla próbki testowej 1,229, dla walidacyjnej 1,061.

#### 3.2. Turbodoładowanie

Dokładność predykcji turbodoładowania za pomocą regresji logistycznej wyniosła 85,08% na próbce treningowej, 86,88% na próbce testowej i 84,29% na próbce walidacyjnej.

Precyzja predykcji turbodoładowania za pomocą regresji logistycznej wyniosła 88% dla wartości pozytywnej i 79% dla wartości negatywnej (na próbce treningowej), 95% dla wartości pozytywnej i 53% dla wartości negatywnej (na próbce testowej) i 87% dla wartości pozytywnej i 80% dla wartości negatywnej (na próbce walidacyjnej)

Czułość predykcji turbodoładowania została zaprezentowana w tablicach pomyłek 3.2.1.:



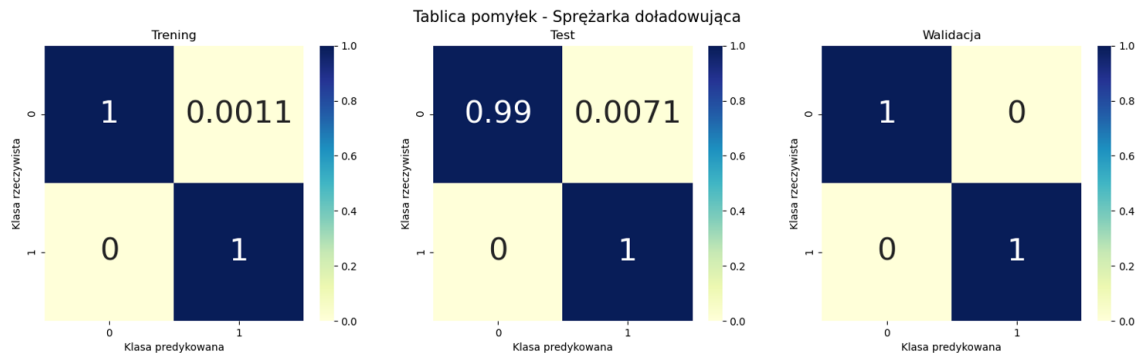
3.2.1. Tablica pomyłek dla turbodoładowania

### 3.3. Sprężarka doładowująca

Dokładność predykcji sprężarki doładowującej za pomocą regresji logistycznej wyniosła 98.89% na próbce treningowej, 99,31% na próbce testowej i 100% na próbce walidacyjnej.

Precyzja predykcji sprężarki doładowującej za pomocą regresji logistycznej wyniosła 100% dla wartości pozytywnej i 94% dla wartości negatywnej (na próbce treningowej), 100% dla wartości pozytywnej i 75% dla wartości negatywnej (na próbce testowej) i 100% dla wartości pozytywnej i 100% dla wartości negatywnej (na próbce walidacyjnej)

Czułość predykcji sprężarki doładowującej została zaprezentowana w tablicach pomyłek 3.3.1.:



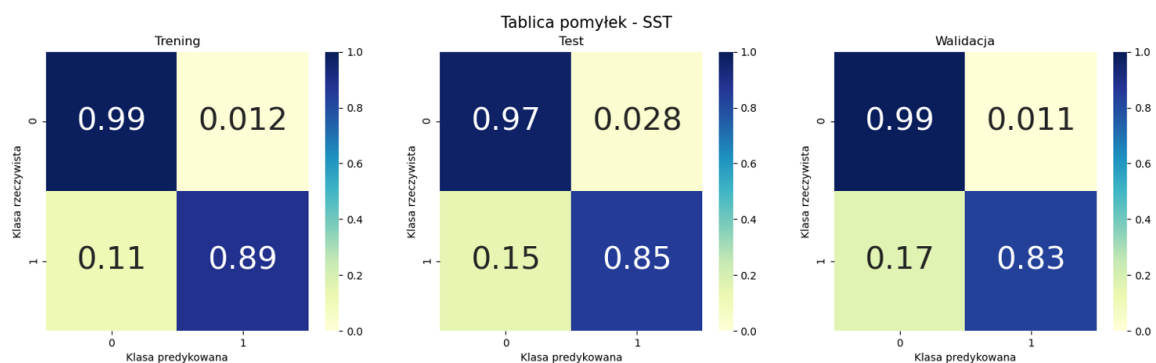
3.3.1. Tablica pomyłek dla sprężarki doładowującej

### 3.4. System Start Stop

Dokładność predykcji SST za pomocą regresji logistycznej wyniosła 98,01% na próbce treningowej, 96,29% na próbce testowej i 97,03% na próbce walidacyjnej.

Precyzja predykcji SST za pomocą regresji logistycznej wyniosła 99% dla wartości pozytywnej i 86% dla wartości negatywnej (na próbce treningowej), 99% dla wartości pozytywnej i 72% dla wartości negatywnej (na próbce testowej) i 98% dla wartości pozytywnej i 91% dla wartości negatywnej (na próbce walidacyjnej)

Czułość predykcji systemu Start Stop została zaprezentowana w tablicach pomyłek 3.4.1.:



3.4.1. Tablica pomyłek dla SST

## Podsumowanie

Zastosowanie modeli uczenia maszynowego pozwalają w prosty sposób rozwiązać skomplikowane zagadnienia. Nie potrzeba dużego nakładu pracy oraz czasu, a rezultaty są obiecujące. Predykcja zużycia paliwa za pomocą tak małej liczby podstawowych cech silnika pozwala od początku projektu silnika mieć bardzo dobrą estymację, w jakiej normie spalania EURO silnik się znajduje. Znając jakie w przybliżeniu silnik ma zużycie paliwa, można za pomocą modelu regresji logistycznej przewidzieć czy nasz silnik potrzebuje konkretne technologie takie jak turbosprężarka, sprężarka doładowująca czy system Start Stop. Przy SST i sprężarce doładowującej model ma dużą dokładność, jest bardzo precyzyjny i ma dużą czułość. Przy turbodoładowaniu dokładność i czułość jest duża, lecz precyzja przy próbie testowej jest niska, zatem model turbodoładowania wymaga dalszych poprawek. Podsumowując powyższa praca ukazuje spore możliwości modeli uczenia maszynowego w inżynierii.

Inne przykładowe zastosowania modeli uczenia maszynowego w inżynierii to na przykład bardzo szybkie tworzenie symulacji mechaniki płynów [7]. Kolejnym zastosowaniem jest kategoryzacja uszkodzeń w silnikach na podstawie zdjęć. Jednak głównym zastosowaniem użycia modeli głębokiego uczenia oraz uczenia maszynowego jest dział sterowania i automatyki. Dzięki zastosowaniu ich można w znaczny sposób podwyższyć bezpieczeństwo i niezawodność w lotu.

## Literatura

- [1] <https://github.com/RusieckiFilip/PracaPrzejsciowa>
- [2] Practical guide to logistic regression - Joseph M. Hilbe
- [3] A Hands-On Introduction to Data Science - Chirag Shah
- [4] Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow - Aurelien Geron
- [5] <https://www.fueleconomy.gov/feg/epadata/21data.zip>
- [6] <https://archive.ics.uci.edu/ml/datasets/auto+mpg>
- [7] Learning to Simulate Complex Physics with Graph Networks - Alvaro Sanchez-Gonzalez
- [8] Imbalanced Learning Foundations, Algorithms, and Applications – Haibo He
- [9] Hands-On Machine Learning with scikit-learn and Scientific Python Toolkits A practical guide to implementing supervised and unsupervised machine learning algorithms in Python - Tarek Amr
- [10] Mastering Matplotlib - Duncan M. McGregor