

Opracowanie do projektu z uczenia maszynowego

Jakub Wojciechowski, Tymoteusz Tomczak
Politechnika Poznańska

Wprowadzenie do sztucznej inteligencji - projekt zaliczeniowy

28.01.2025

1 Omówienie metod wykrywania usterek w silnikach przemysłowych

1.1 Pierwsza metoda: Analiza na podstawie regresji liniowej oraz Random Forest ¹

Pierwszą metodą analizy danych temperaturowych jest wykorzystanie regresji liniowej jako modelu bazowego oraz algorytmu Random Forest, który lepiej uchwycił nieliniowe zależności w danych, przewyższając dokładnością regresję liniową.

Mocne strony:

- Random Forest doskonale modeluje nieliniowości i działa efektywnie nawet w przypadku większej liczby cech.
- Możliwość interpretacji wyników dzięki funkcji oceny ważności cech.

Słabe strony:

- Wyższe wymagania obliczeniowe w porównaniu do prostszych modeli, takich jak regresja liniowa.
- Random Forest może przeuczać się na małych lub nierównoważnych zbiorach danych.

1.2 Druga metoda: Klasyfikatory K-NN i SVM²

W pracy „A Machine Learning Framework for Bearing Fault Detection in Three-Phase Induction Motors” zaproponowano wykorzystanie klasyfikatorów K-Nearest Neighbors (K-NN) oraz Support Vector Machine (SVM) do przewidywania temperatury w silnikach PMSM. Porównano skuteczność obu algorytmów, wskazując, że:

¹S. Sawant, “Electrical Motor Temperature: PMSM,” 2021. Dostępne online: kaggle.com.

²W. Rohouma, A. Zaitouny, M. F. Wahid, H. Ali and S. S. Refaat, “A Machine Learning Framework for Bearing Fault Detection in Three-Phase Induction Motors,” 2024 4th International Conference on Smart Grid and Renewable Energy (SGRE), Doha, Qatar, 2024, pp. 1-5, doi: 10.1109/SGRE59715.2024.10429024.

- K-NN dobrze działa przy lokalnych zależnościach w danych.
- SVM wykazuje lepszą zdolność generalizacji w przypadku bardziej złożonych zestawów danych.

Mocne strony:

- **K-NN:** Prosty w implementacji, nie wymaga długiego procesu trenowania i dobrze radzi sobie z lokalnymi wzorcami w danych.
- **SVM:** Wysoka dokładność predykcji, szczególnie w przypadku złożonych problemów, dzięki wykorzystaniu nieliniowych funkcji jądrowych.

Słabe strony:

- **K-NN:** Wysoka złożoność obliczeniowa w czasie predykcji, szczególnie dla dużych zbiorów danych.
- **SVM:** Wymaga starannego doboru parametrów (np. funkcji jądrowej) i jest mniej efektywny przy bardzo dużych zbiorach danych.

1.3 Trzecia metoda: Thermal Imaging z wykorzystaniem BNN ³

W pracy „Diagnosis of the three-phase induction motor using thermal imaging” autorzy zaproponowali autorskie rozwiązanie wykorzystujące Bayesowskie Sieci Neuronowe (BNN), klasyfikator najbliższych sąsiadów oraz algorytm centroidów.

Cechy BNN:

Mocne strony:

- Modelowanie niepewności: Podejście probabilistyczne pozwala oszacować pewność wyników.
- Elastyczność: Możliwość modelowania skomplikowanych zależności w danych.
- Solidność: Mniejsza podatność na przeuczenie w porównaniu do klasycznych sieci neuronowych.
- Interpretowalność: Wagi z rozkładami probabilistycznymi dostarczają dodatkowych informacji o znaczeniu cech.

Słabe strony:

- Wysoka złożoność obliczeniowa (np. wymagane techniki Monte Carlo).
- Wysokie wymagania dotyczące danych w celu dokładnego oszacowania rozkładów wag.
- Trudności w skalowaniu na bardzo dużych zbiorach danych.

³Adam Glowacz, Zygfryd Glowacz, Diagnosis of the three-phase induction motor using thermal imaging, Infrared Physics Technology, Volume 81, 2017, Pages 7-16, ISSN 1350-4495, 10.1016/j.infrared.2016.12.003 www.sciencedirect.com

Cechy algorytmu centroidów:

Mocne strony:

- Szybkość działania i prostota implementacji.
- Efektywność dla dobrze rozdzielnych danych.

Słabe strony:

- Niska dokładność w przypadku złożonych zbiorów danych.
- Brak adaptacyjności i wrażliwość na rozkład danych.

2 Kontekst pracy oraz założenie dotyczące pracy programu

Założeniem programu było przetestowanie 3 metod oraz sprawdzenie, który algorytm i w jakich warunkach najlepiej sobie radzi. Celem było określenie, czy model na podstawie zmiennych procesowych jest w stanie przewidzieć przegrzanie silnika, które mogłoby prowadzić do jego uszkodzenia.

2.1 Baza danych

Dane użyte w projekcie pochodzą ze strony Kaggle i zostały zebrane przez naukowców z Uniwersytetu w Paderborn. Zostały zmniejszone do 17000 rekordów i zawierały 12 kolumn:

1. ID
2. U_q – składowa q napięcia
3. Coolant – temperatura materiału chłodzącego
4. Stator_winding – temperatura uzwojenia stojana
5. U_d – składowa d napięcia
6. Stator_tooth – temperatura zębów silnika
7. Motor_speed – prędkość obrotowa silnika
8. I_d – składowa d prądu
9. I_q – składowa q prądu
10. Pm – temperatura magnesów trwałych stojana
11. Stator_yoke – temperatura obudowy
12. Torque – moment obrotowy
13. Profile_id – oznaczenie sesji pomiarowej

Dla naszego problemu modele otrzymają wszystkie pola oprócz pól temperaturowych oraz Profileid (nie wpływa to na wyniki pomiarowe, ponieważ do nauki korzystamy tylko i wyłącznie z jednego Id). Fabrycznie modele przewidują temperaturę magnesów. Odpowiednio dla metody regresji logistycznej dane zostały ustandaryzowane metodą `StandardScaler()`, a dane do SVM zostały znormalizowane metodą `MinMaxScaler()`.

Program po kilku modyfikacjach jest też gotowy do przewidywania innych zmiennych procesowych, wystarczy zamienić kolumny które nie są brane pod uwagę podczas fazy treningowej oraz co jest zmienną do przewidzenia (zmienna y). Pod koniec pracy programu wyświetlane są wykresy drzewa decyzyjnego, dokładności każdej z trzech metod, macierze konfuzji oraz macierze korelacji.

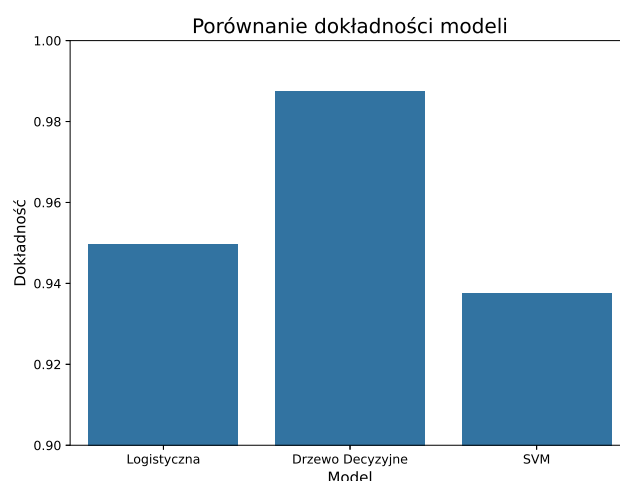
2.2 Przygotowanie danych

- Dla regresji logistycznej dane zostały ustandaryzowane za pomocą `StandardScaler()`.
- Dla SVM zastosowano normalizację metodą `MinMaxScaler()`.

Program umożliwia również przewidywanie innych zmiennych procesowych poprzez zmianę kolumn ignorowanych podczas treningu oraz zmiennej przewidywanej (y).

3 Przedstawienie wyników

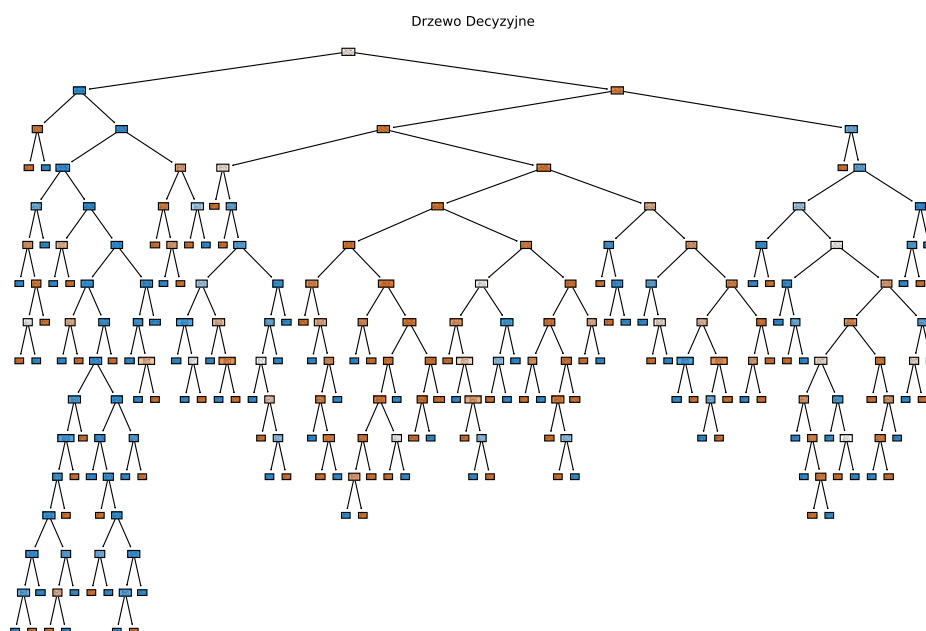
3.1 Wyniki dokładności:



Rysunek 1: Wykresy dokładności modeli

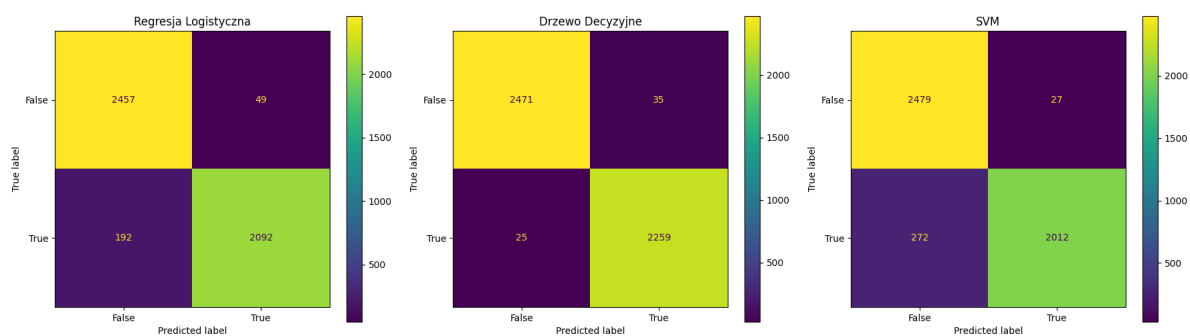
- **Drzewo decyzyjne:** Dokładność 98,7%.
- **Regresja logistyczna:** Dokładność 94,9%.
- **SVM:** Dokładność 93,8%.

3.2 Drzewo decyzyjne



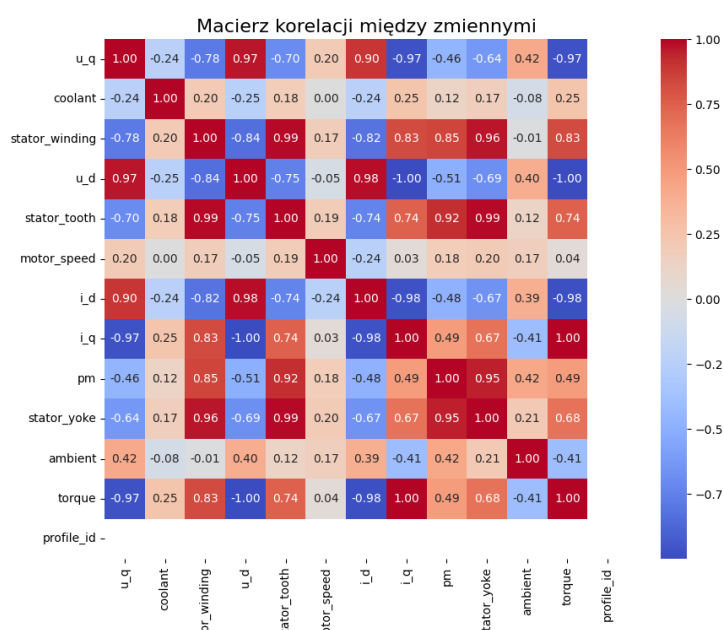
Rysunek 2: Drzewo decyzyjne

3.3 Macierze konfuzji

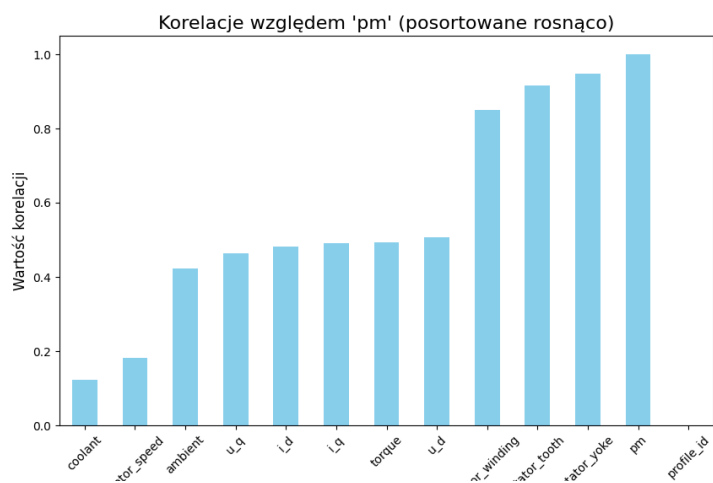


Rysunek 3: Macierze konfuzji

3.4 Korelacja



Rysunek 4: Macierz korelacji



Rysunek 5: Przedstawienie korelacji rosnąco

4 Wnioski

Wszystkie modele doszły do bardzo zadowalających poziomów dokładności. Drzewo decyzyjne, pomimo najwyższej dokładności nie wydaje się być adekwatne do tego typu problemów, ponieważ staje się bardzo rozległe i ciężkie do analizy. Nie można też jednoznacznie stwierdzić, który z pozostałych modeli będzie zachowywać się lepiej, ponieważ nasza próba opierała się na mniejszej ilości próbek niż w przypadku przemysłowym, oraz obydwa modele zachowują bardzo podobną celność przewidywania. Na podstawie macierzy korelacji można dostrzec, że wszystkie temperatury są wprost proporcjonalne do siebie, mają bardzo duże wskaźniki na poziomie prawie jedności. Zatem usunięcie jakiegokolwiek pomiaru temperatury powoduje zwiększenie poziomu trudności problemu dla modeli. Najmniej użytecznym pomiarem wydaje się być prędkość obrotowa, ponieważ korelacja z żadną zmienną nie przekracza 0,2. W kontekście zastosowania takiego rozwiązania w przemyśle należałoby zapewnić bezpośrednio pomiary napięcia, prądu i momentu obrotowego. Jeżeli istniałaby możliwość odczytania jakiegokolwiek parametru temperaturowego, wtedy dokładność modeli bardzo wzrośnie, właśnie z uwagi na korelację. Chcąc zastosować takie rozwiązanie w zakładach wykorzystujących tego typu silniki, należy zwrócić uwagę na trudność modyfikacji istniejących już układów. Problematicznym mogłoby być przystosowanie stanowiska pomiarowego do tak dokładnego i częstego pomiaru wielu wielkości procesowych, np. momentu obrotowego. Najbardziej rozsądnym wyborem wydaje się wykorzystywać tego typu pomysły w nowych projektach, gdzie od początku założeniem będzie akwizycja danych, po to aby móc przewidywać potencjalne usterki. W takiej sytuacji według nas wymagany jest pomiar napięcia, prądu, momentu obrotowego oraz chociaż jednego parametru temperaturowego - np. temperatury stojana. Można to zrealizować w bardzo prosty sposób za pomocą czujnika temperatury, a może to bardzo ułatwić predykcję w przypadku wielu, długotrwale zbieranych próbek. Jednocześnie widzimy zastosowanie, w którym predykcji nie będzie ulegać pomiar magnesów trwałych, tylko na przykład części podlegające tarcia - łożyska, które też ulegając zwiększonej temperaturze są przyczyną usterek, a co za tym idzie strat finansowych w przemyśle.