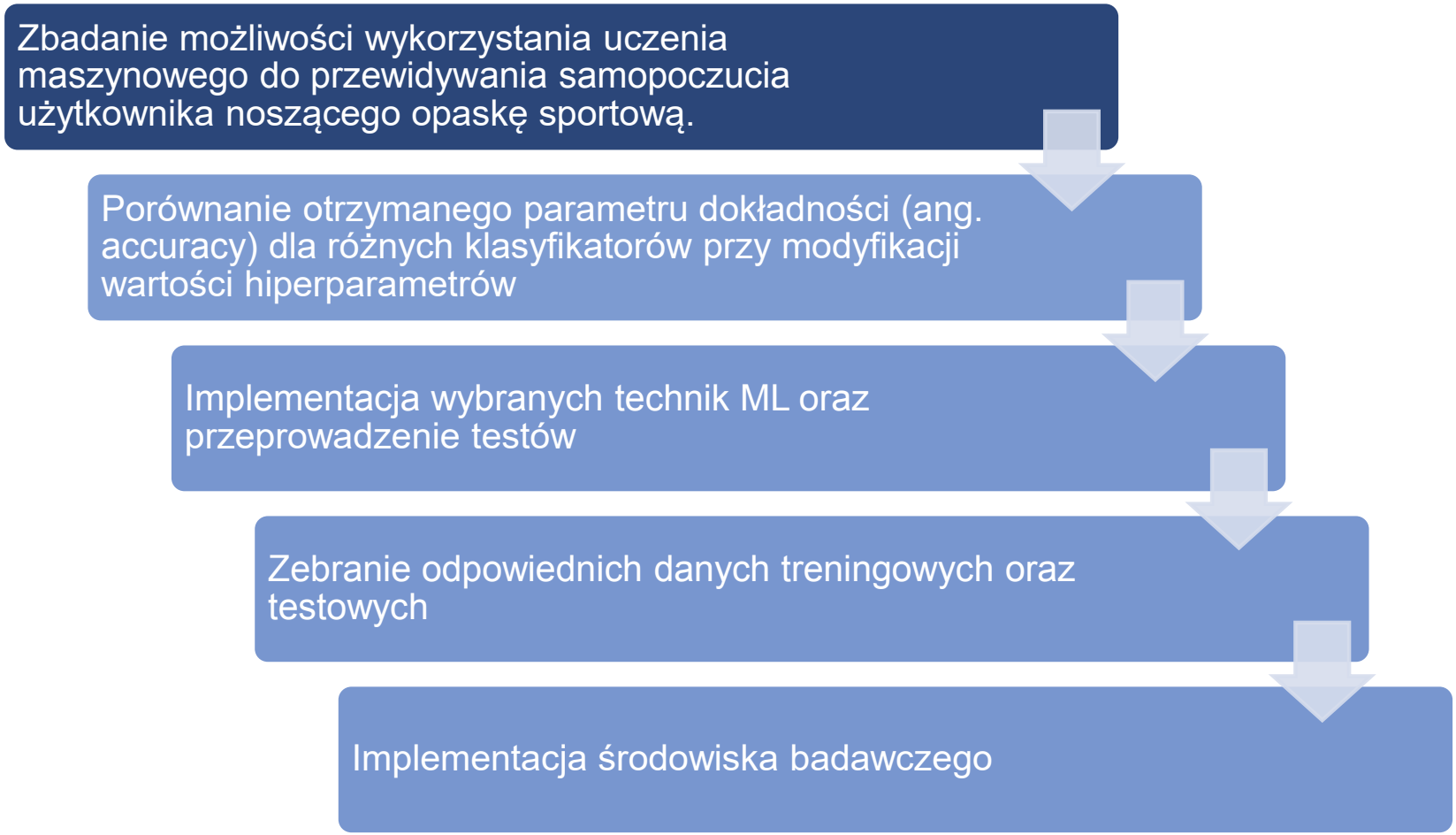
Politechnika Warszawska

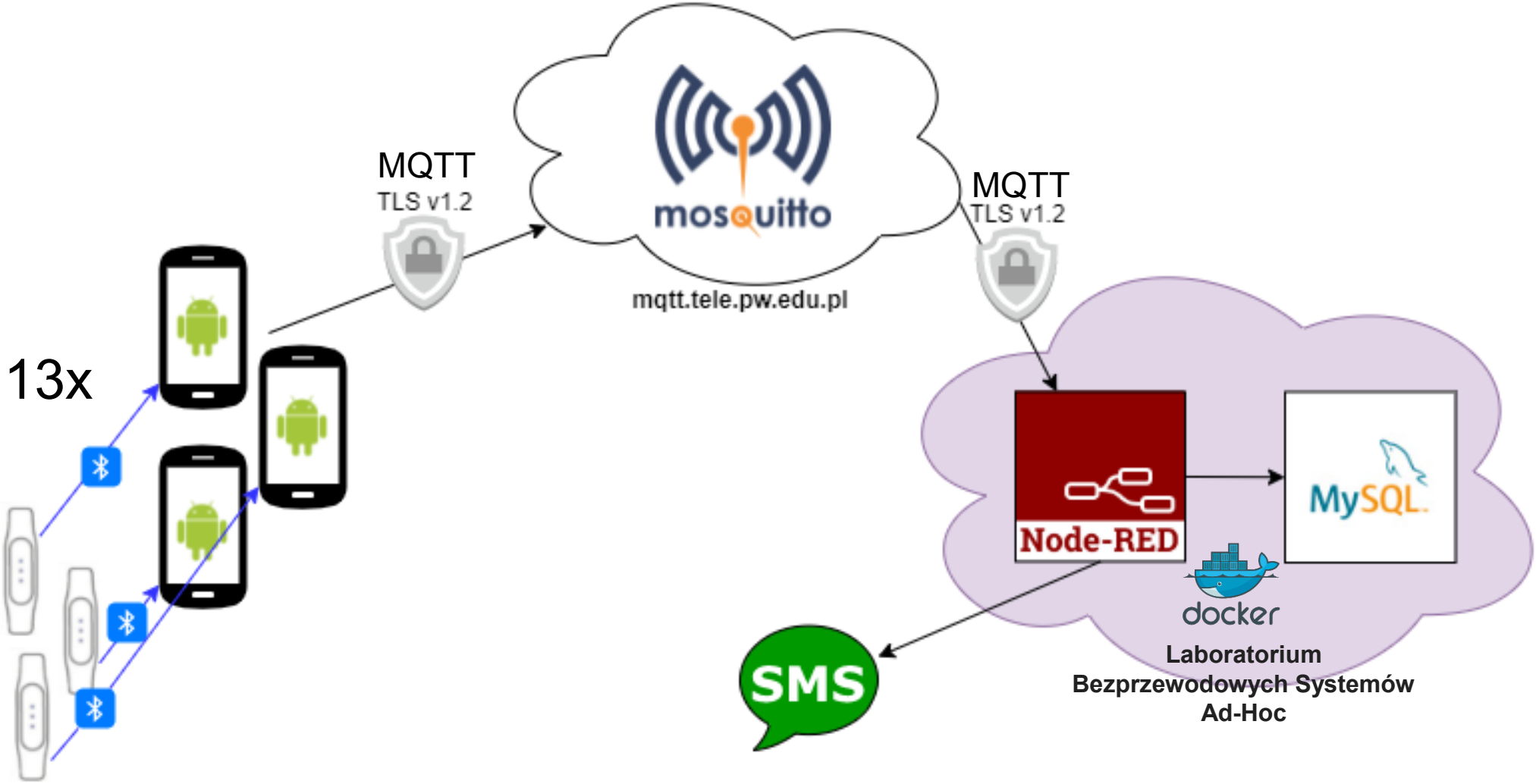
Prognozowanie samopoczucia z wykorzystaniem algorytmów uczenia maszynowego dla sieci opasek sportowych.

Autor: Beata Kogut
Promotor: dr hab. inż. Krzysztof Perlicki, prof. uczelni
Opiekun pracy: mgr inż. Marcin Golański

Cel pracy


Zbadanie możliwości wykorzystania uczenia maszynowego do przewidywania samopoczucia użytkownika noszącego opaskę sportową.
Porównanie otrzymanego parametru dokładności (ang. accuracy) dla różnych klasyfikatorów przy modyfikacji wartości hiperparametrów
Implementacja wybranych technik ML oraz przeprowadzenie testów
Zebranie odpowiednich danych treningowych oraz testowych
Implementacja środowiska badawczego

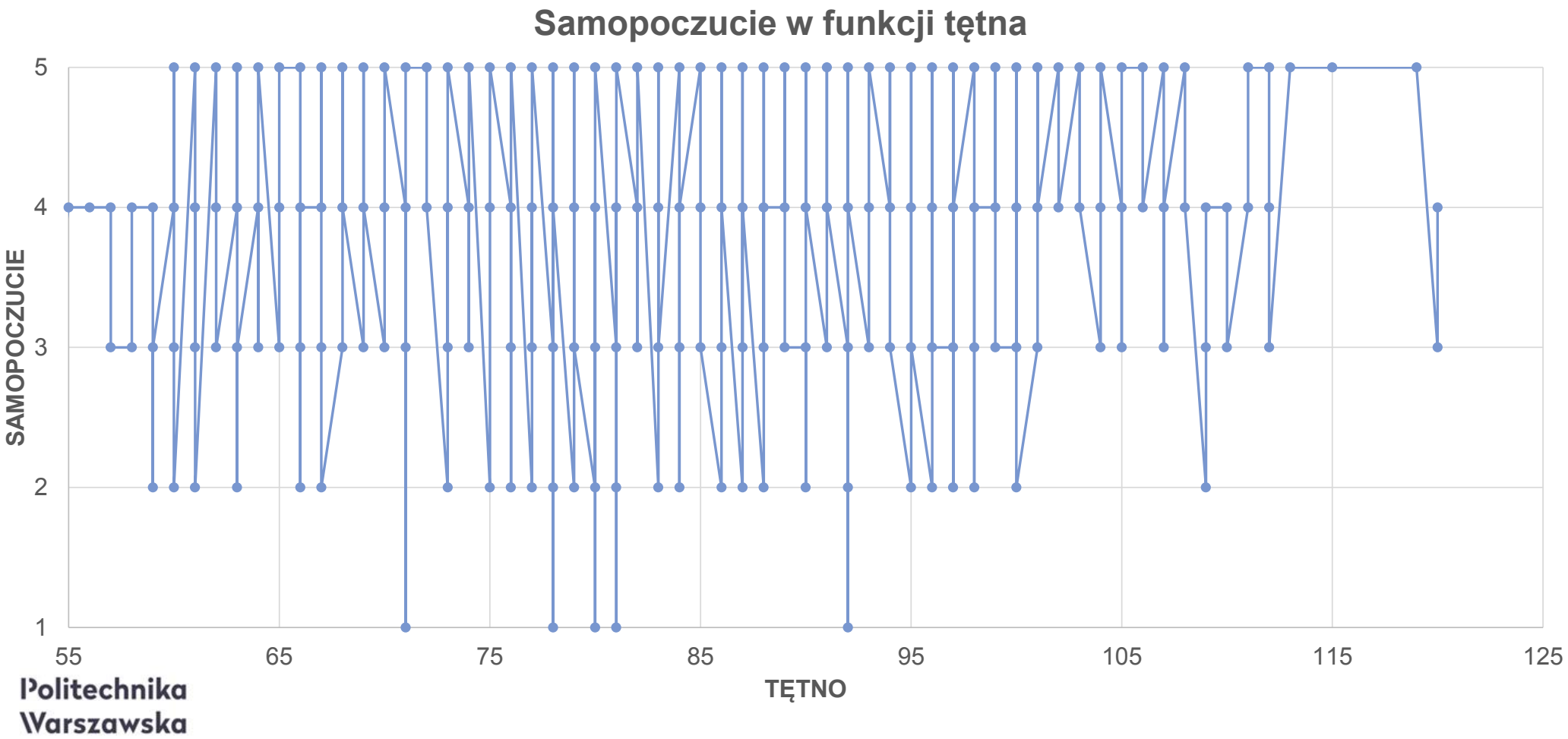
Architektura rozwiązania



Proces zbierania danych

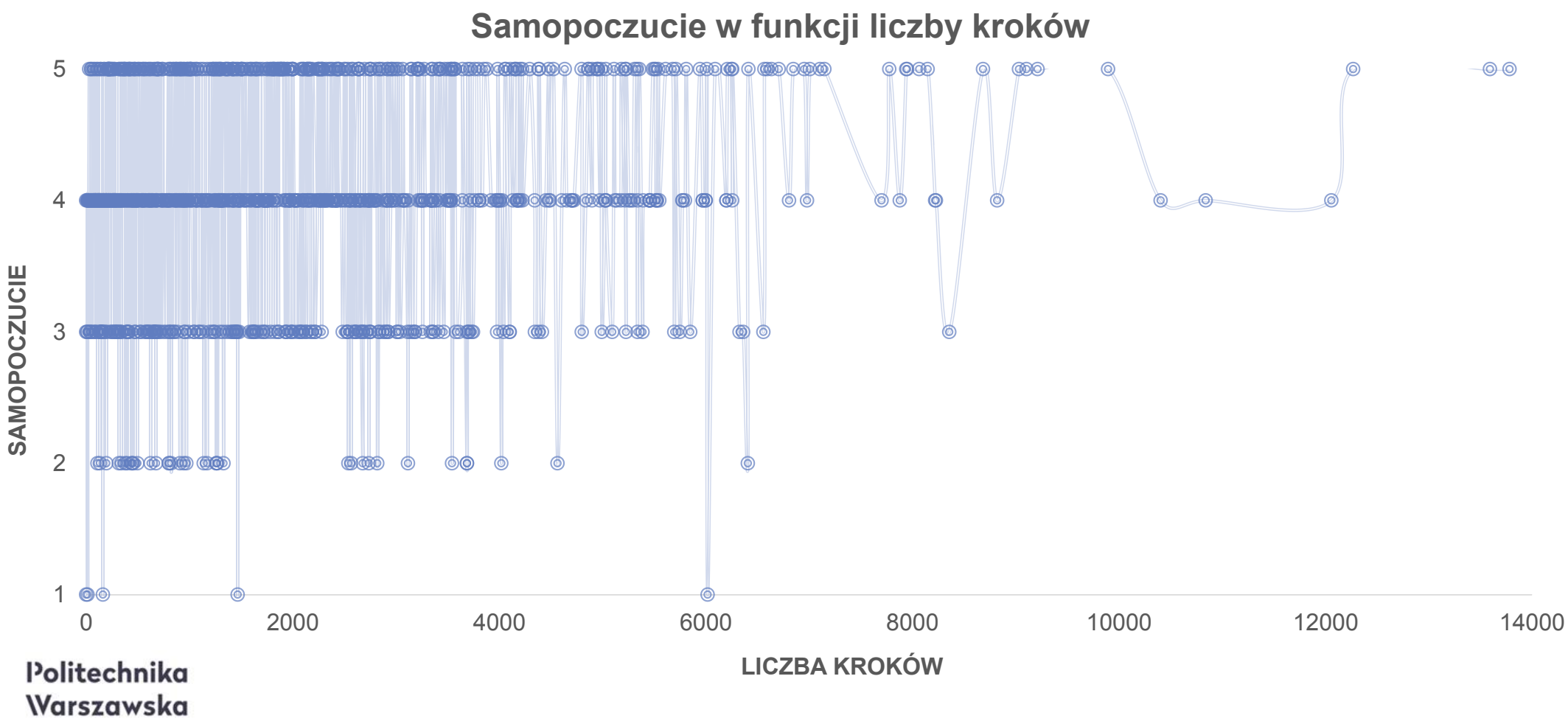
1. Zbierane informacje:
   - Liczba kroków
   - Tętno
     - Pomiar co 5 minut
   - Samopoczucie
     - Skala ocen od 1 do 5
     - Pomiar samopoczucia 3 razy dziennie o określonej porze dnia
     - Ankieta w aplikacji
   - Znacznik czasowy
     - Każda zebrana wartość oznaczona jest znacznikiem czasowym
   - Id urządzenia pomiarowego MiBand
2. Czas zbierania danych
   - Dane treningowe: 30 dni
   - Dane testowe: 3-4 dni
   - Łącznie zebrano ponad 1100 rekordów danych

Analiza danych



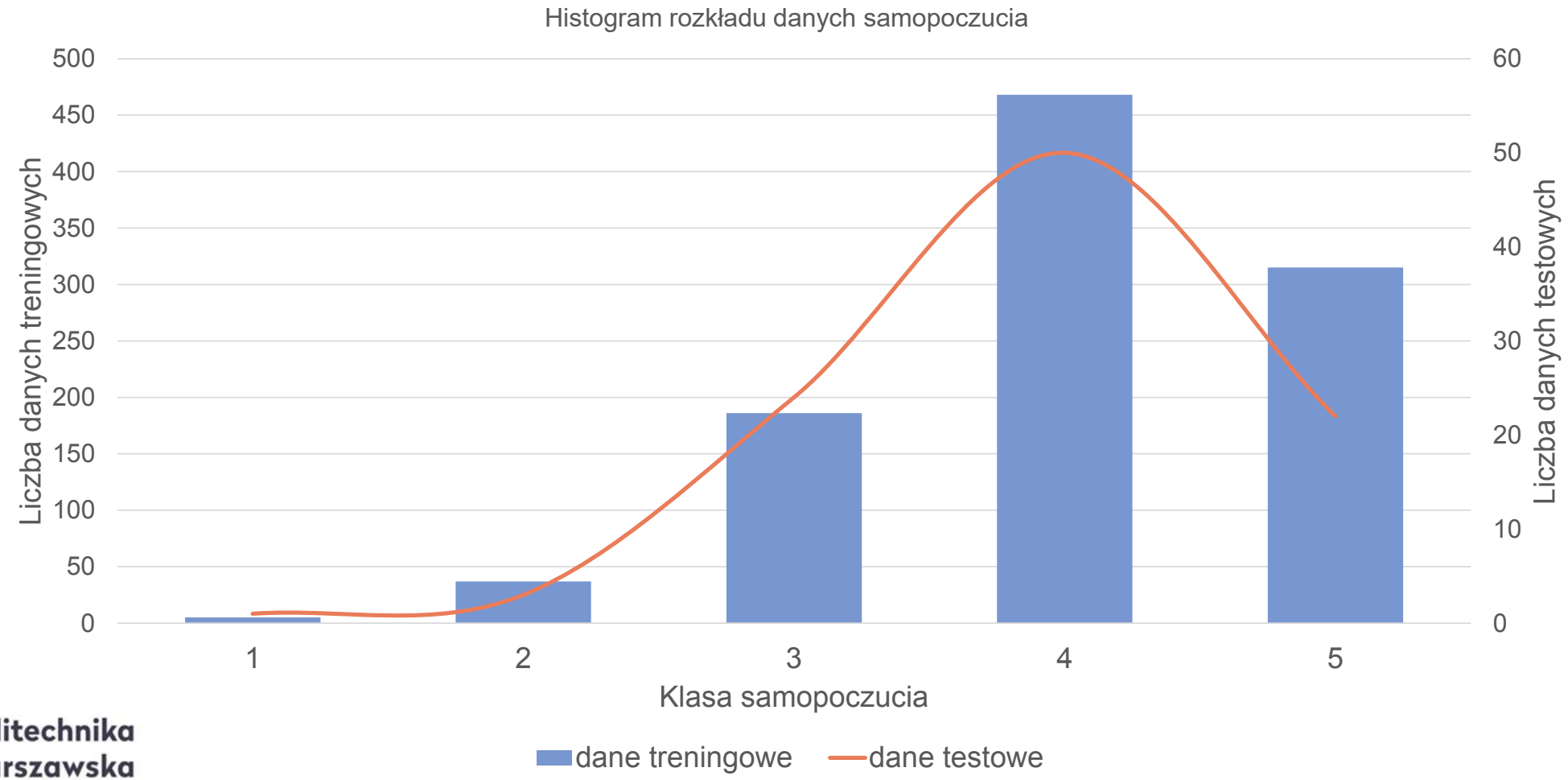
Politechnika Warszawska

Analiza danych



Politechnika Warszawska

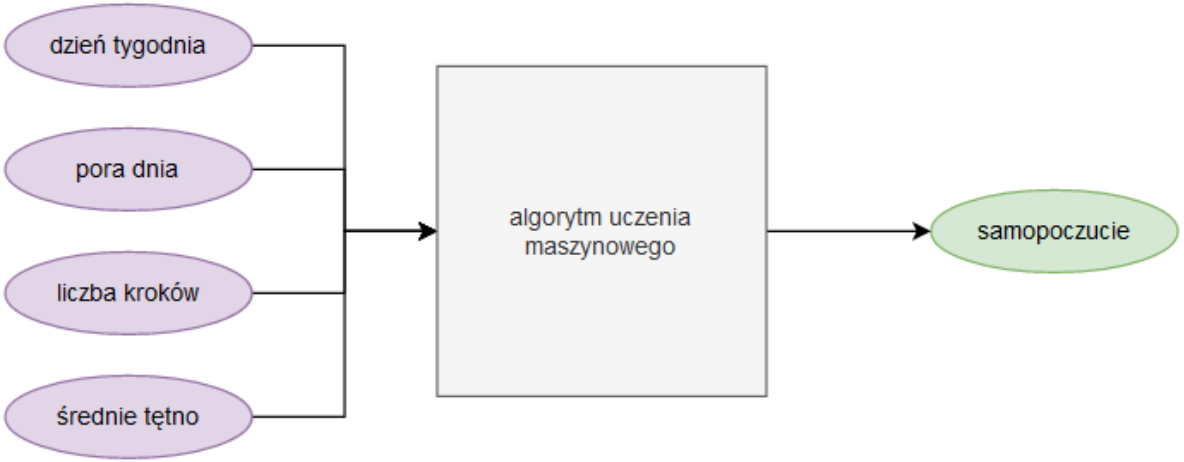
Analiza danych – histogram samopoczucia



Politechnika Warszawska

Klasyfikacja - założenia

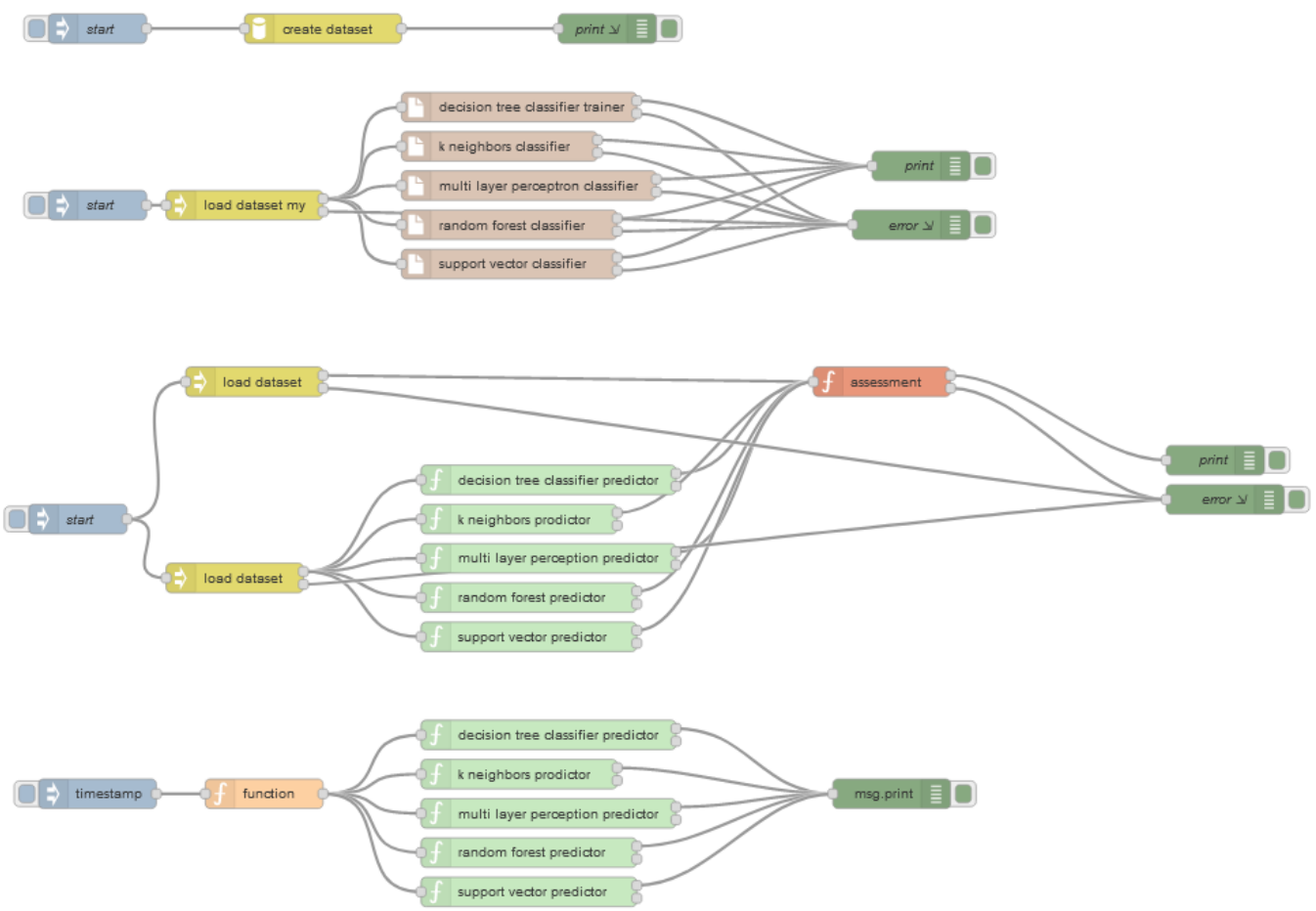
1. Dane wejściowe do ML
   - Dzień tygodnia
   - Pora dnia
   - Liczba kroków
   - Średnia wartość tętna
2. Dane wyjściowe
   - Samopoczucie użytkownika (1:5)



Klasyfikacja – wybrane algorytmy

- Drzewo decyzyjne (Decision Tree Classifier)
- K-najbliższych sąsiadów (K-nearest Neighbors)
- Klasyfikator wielowarstwowy (Multilayer Perception Classifier)
- Losowy las decyzyjny (Random Forest Classifier)
- Klasyfikator wektora nośnych (Support Vector Classifier)

Klasyfikacja – implementacja w środowisku Node-RED

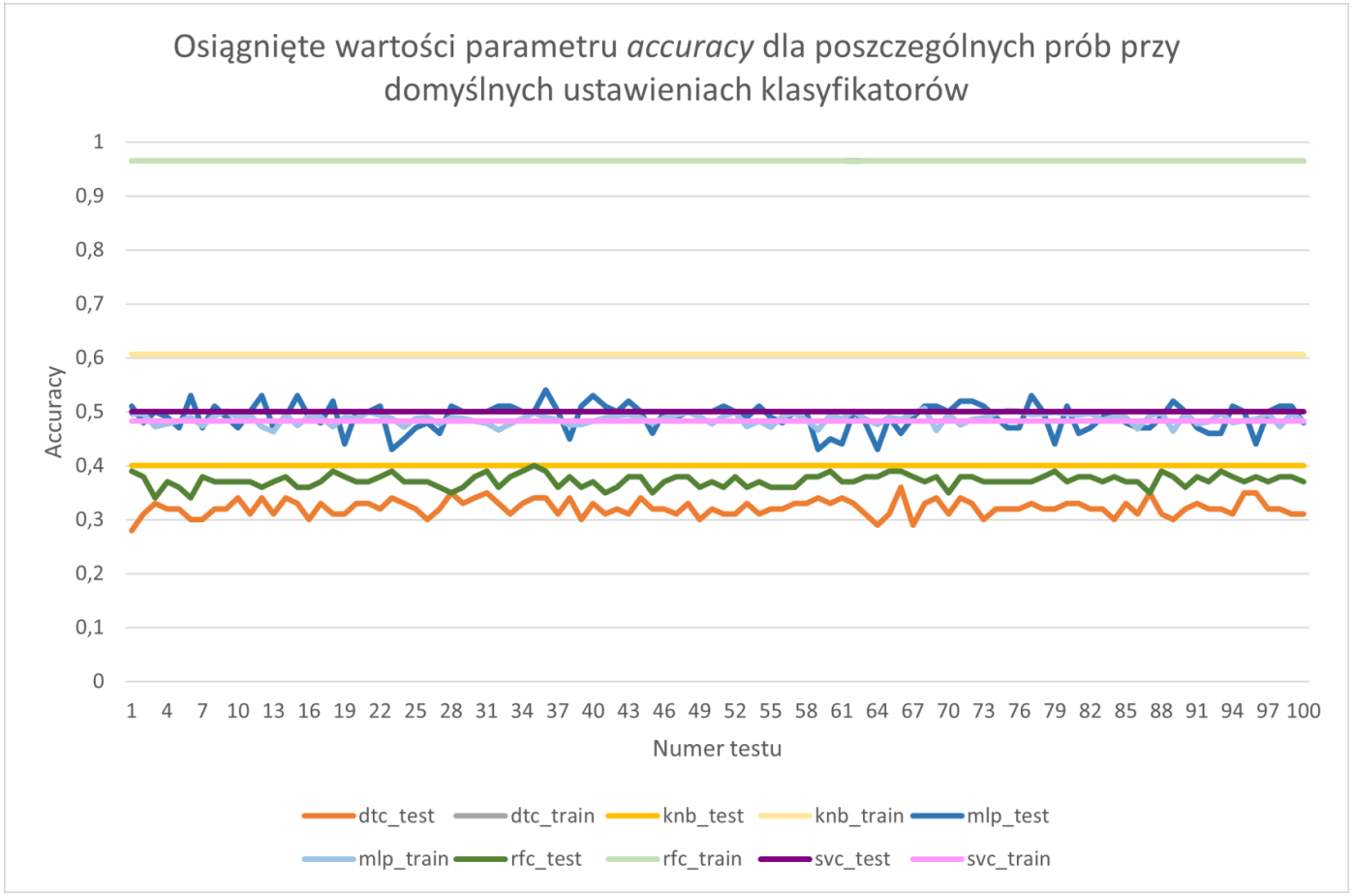


1. Stworzenie zbiorów danych
   - Dane treningowe
   - Dane testowe
   - Format csv
2. Utworzenie modelu dla każdego klasyfikatora na podstawie danych treningowych
3. Wczytanie danych testowych i przetestowanie modelu

   Dokonanie oceny modelu za pomocą parametru dokładności (ang. accuracy)
4. Przewidywanie wyniku za pomocą wybranego modelu

Eksperyment 1 – domyślne ustawienia

1. Badanie na domyślnych ustawieniach klasyfikatorów
2. Problem nadmiernego dopasowania (ang. overfitting) modelu
3. Wyniki parametru *accuracy* na niskim poziomie
   ok. 40% - max 55%



Eksperyment 2 – optymalizacja hiperparametrów

1. Wzrost wartości parametru *accuracy* dla **każdego** z badanych klasyfikatorów
2. Algorytm drzewa decyzyjnego najbardziej reagował na zmianę parametrów
3. Wzrost dokładności o max. 20 p.p

	Ustawienia domyślne	Zoptymalizowane parametry (średnia)	Zoptymalizowane parametry (max)
Decision tree	32%	38% (Wzrost o 18,75% ->6 p.p.)	52% Wzrost o 62,5% ->20 p.p)
K neighbors	40%	45% (Wzrost o 12,5 % ->5 p.p.)	45% (Wzrost o 12,5% ->5.p.p.)
Multilayer perception	49%	52% (Wzrost o 6,12% ->3 p.p.)	56% (Wzrost o 14,28% ->7 p.p.)
Random forest	37%	40% (Wzrost o 8,1% ->3 p.p.)	42% (Wzrost o 13,51% ->5 p.p.)
Support vector	50%	52% (Wzrost o 4% ->2p.p.)	52% (Wzrost o 4% ->2 p.p.)

Tabela 1. Porównanie osiągniętych wartości parametru *accuracy* przed i po optymalizacji

Eksperyment 3 – problem niezbalansowania danych

1. Metody
   - Losowe zwiększanie liczebności mniejszościowych klas (ang. Oversampling minorityclass) - w małych klasach losowo dodajemy kopie danych
   - Dodanie sztucznie wygenerowanych danych do mniejszościowych klas.
     - SMOTE (ang. Synthetic Minority OversamplingTechnique)
2. Wzrost parametru *accuracy* do poziomu prawie 60%
3. Algorytm drzewa decyzyjnego przewidział poprawnie:
   - 0% próbek z klasy 1
   - 100% próbek z klasy 2
   - 41.67% próbek z klasy 3
   - 60% próbek z klasy 4
   - 72.72% próbek z klasy 5

Politechnika Warszawska

Multilayer perception						
przewidywana wartość	1	0	0	0	0	0
	2	0	0	0	0	0
	3	0	0	8	6	0
	4	1	3	14	42	14
	5	0	0	2	2	8
dokładność:58%		1	2	3	4	5
		rzeczywista wartość				

Decision tree						
przewidywana wartość	1	0	0	0	0	1
	2	1	3	1	2	0
	3	0	0	10	10	1
	4	0	0	9	30	4
	5	0	0	4	8	16
dokładność:59%		1	2	3	4	5
		rzeczywista wartość				

Tabela 2. Macierze pomyłek dla klasyfikatorów z najwyższą dokładnością

Eksperyment 4 – przejście na problem binarny

1. Możliwość rozwoju stworzonego rozwiązania: reagowanie systemu w przypadku wykrycia pogorszenia samopoczucia użytkownika
2. Utworzono dwie klasy decyzyjne
    - 0 - złe samopoczucie -> wyślij alert
    - 1 - dobre samopoczucie -> nic nie rób
3. Wyniki tożsame z wynikami osiągniętymi dla pięciu klas decyzyjnych
4. Najlepszym klasyfikatorem dla rozważanego rozwiązania jest klasyfikator drzewa decyzyjnego
5. Dokładność na poziomie 77%

Multilayer perception classifier			
przewidy	0	6	2
wana	1	22	70
		0	1
dokładność:76%		rzeczywista wartość	

Decision tree classifier			
przewidy	0	14	9
wana	1	14	63
		0	1
dokładność:77%		rzeczywista wartość	

Tabela 3. Macierze pomyłek dla najlepszych klasyfikatorów

Wnioski

Możliwe jest wykorzystanie uczenia maszynowego do przewidywania samopoczucia użytkownika noszącego opaskę sportową.

W rozważanym rozwiązaniu najlepsze wyniki uzyskano za pomocą klasyfikatora drzewa decyzyjnego.

Przy pięciu klasach decyzyjnych osiągnięto maksymalną dokładność 59%

Dla problemu binarnego uzyskano dokładność 77%

Możliwości rozwoju i udoskonalenia rozwiązania

Poszerzenie wektora danych wejściowych (nowe parametry np. temperatura ciała, natlenienie krwi)

Profilowanie użytkowników np. ze względu na wiek, płeć, wagę, tryb życia

Zebranie większej liczby danych treningowych, w szczególności danych dla klasy 1

Wprowadzenie korelacji pomiędzy występującymi po sobie pomiarami

Politechnika Warszawska

Dziękuję za uwagę

Politechnika Warszawska

Prognozowanie samopoczucia z wykorzystaniem algorytmów uczenia maszynowego dla sieci opasek sportowych.

Autor: Beata Kogut
Promotor: dr hab. inż. Krzysztof Perlicki, prof. uczelni
Opiekun pracy: mgr inż. Marcin Golański