



**Wydział  
Fizyki**

POLITECHNIKA WARSZAWSKA

# Klasyfikacja typu windy na podstawie danych z czujników komórkowych podczas jazdy przy użyciu algorytmów BigData

**Wykonali:**

Maciej Czarnecki

Denys Morokov

**Politechnika  
Warszawska**



# Cel projektu

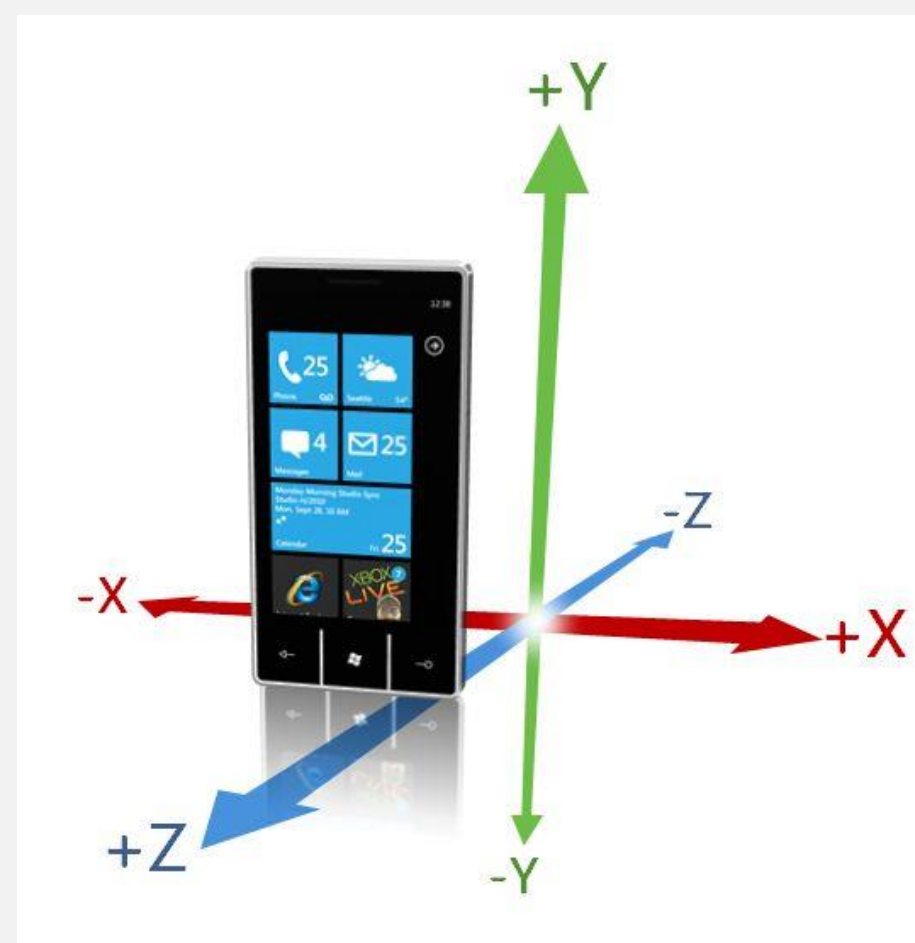
- Na podstawie danych zebranych za pomocą wbudowanych czujników w telefonie komórkowym podczas jazdy windą sklasyfikować odmianę dźwięgu przy użyciu algorytmów BigData.

# Metodyka

- **Dane:** ruch w obie strony 3. wind – Gmach Fizyki, Gmach Główny, MiNI.
- **Akwizycja danych:** phyphox – aplikacja do rejestracji danych z wbudowanych sensorów komórkowych.
- **Wykorzystane sensory:** akcelerometr, żyroskop.
- **Język programowania:** R
- **Modele telefonów:** iPhone 6, Honor 10 Lite
- Dane zbierano na 2. telefony komórkowe – zbiór uczący i zbiór testowy.

# Analiza zebranych danych

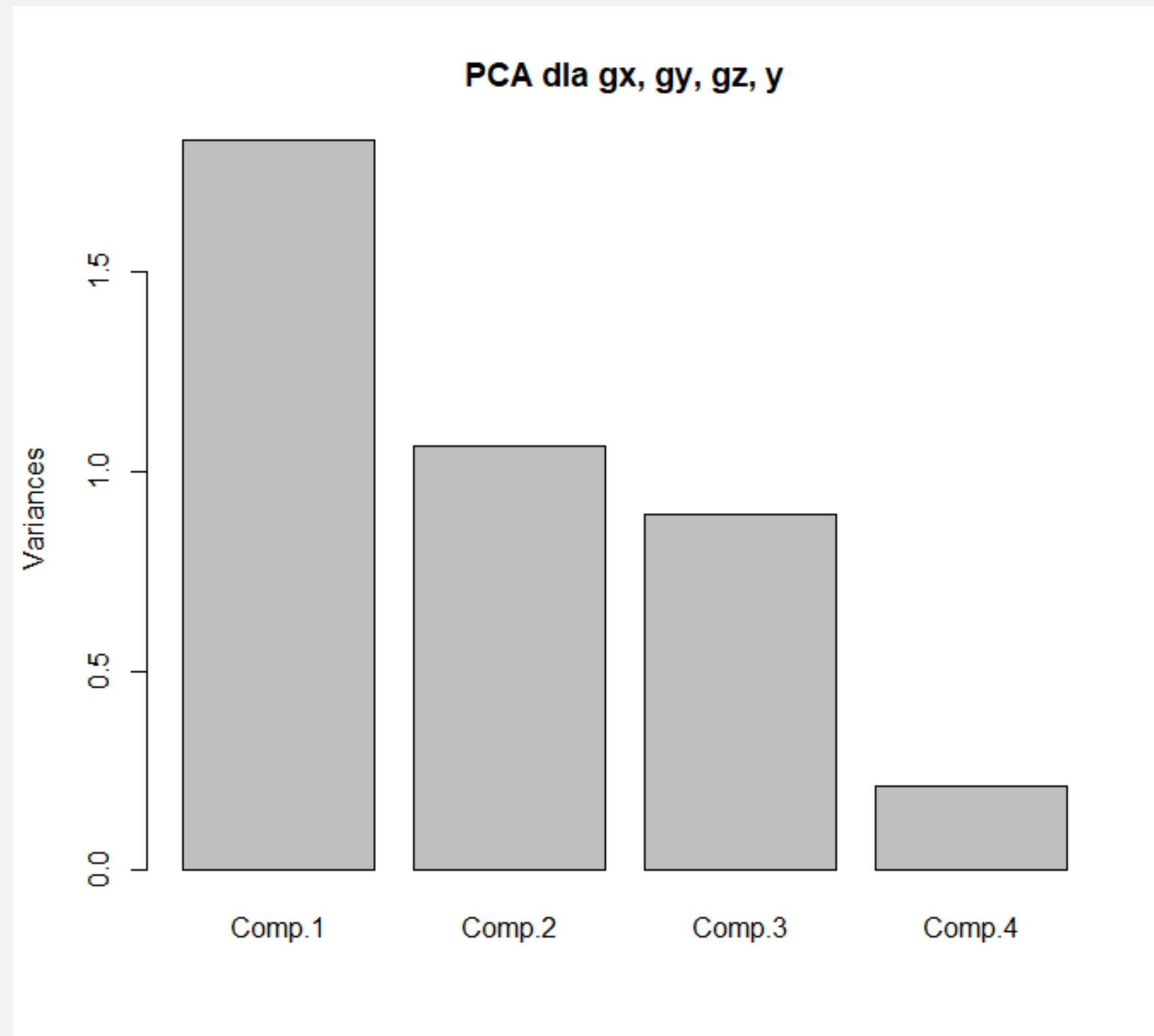
- Uzyskano dane z 3 składowych żyroskopu oraz akcelometru, ale z akcelometru użyto do klasyfikacji tylko składowa Y.
- W celu określenia, które z otrzymanych składowych są istotne użyto metodę **analizy składowych głównych (PCA)**.



Rys. 1. Składowe akcelometru [1]

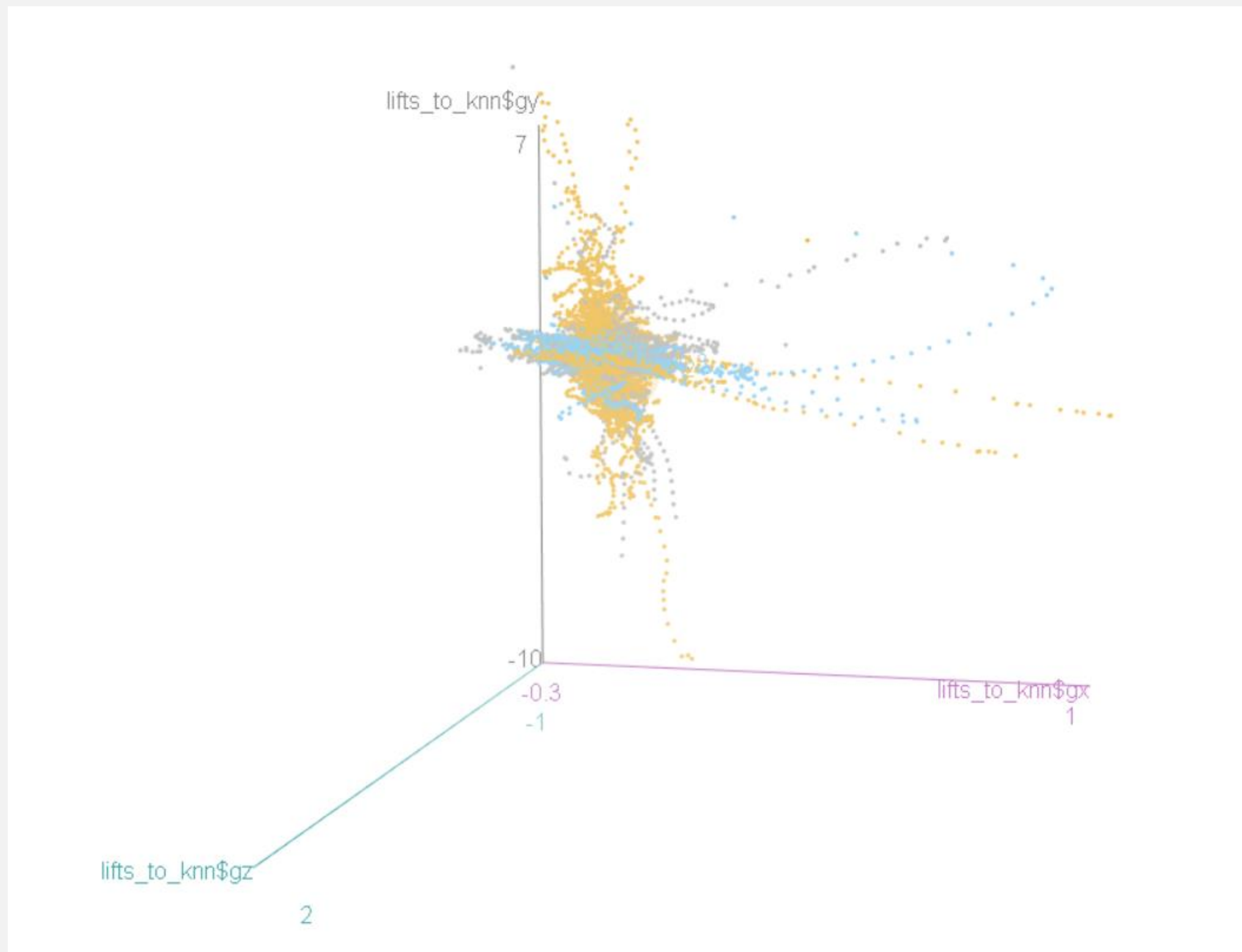
# PCA

- Bezpośrednio do uczenia klasyfikatora użyto dane wyłącznie z żyroskopu.

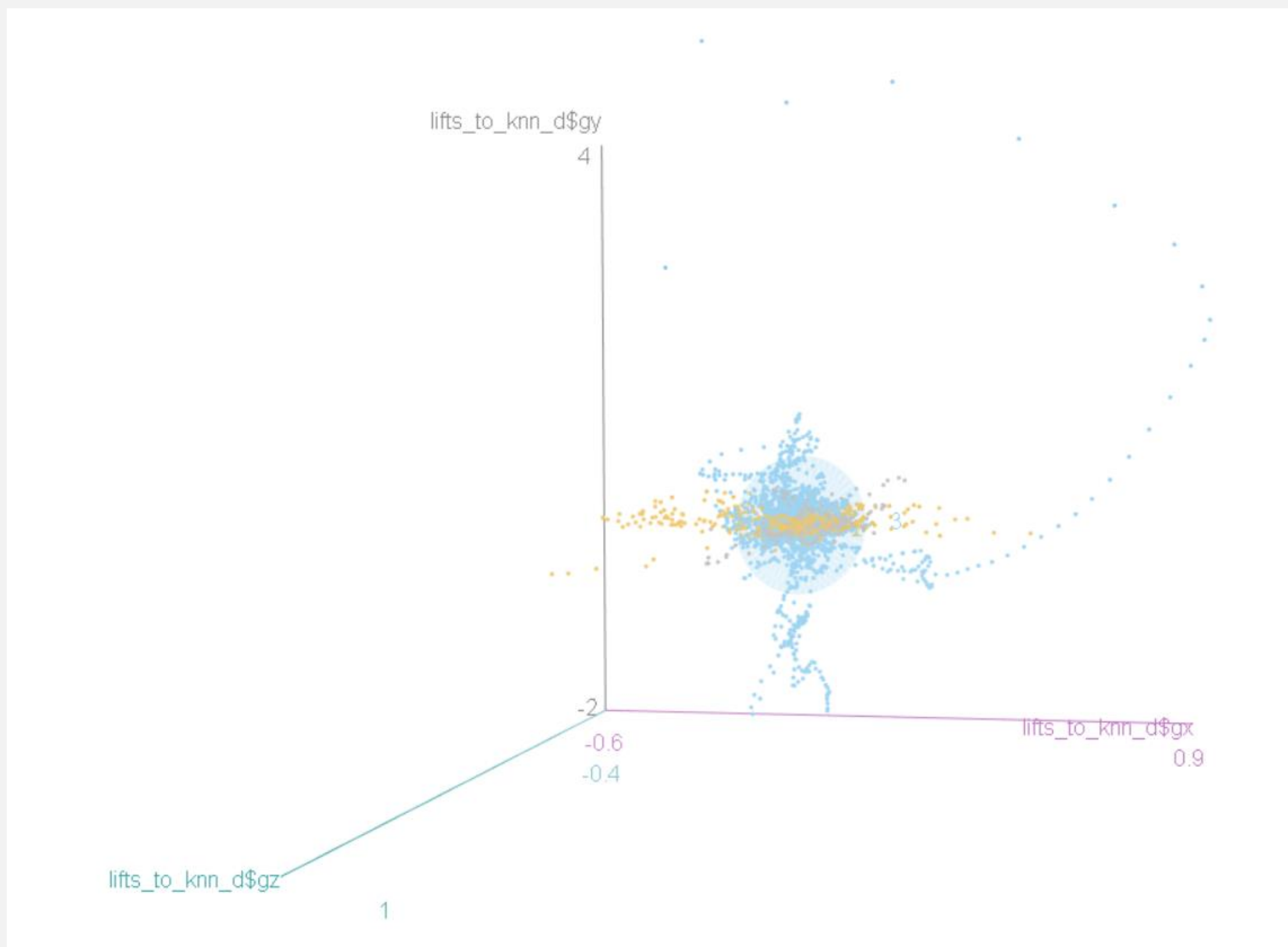


Rys. 2. Wyniki PCA dla 3. składowych z żyroskopu i 1. z akcelerometru

# Wizualizacja danych



Rys. 3. Dane z żyroskopu, zbiór uczący.  
Szary – GF, żółty – MiNI, niebieski – GF



Rys. 4. Dane z żyroskopu, zbiór testowy.  
Szary – GF, żółty – MiNI, niebieski – GF

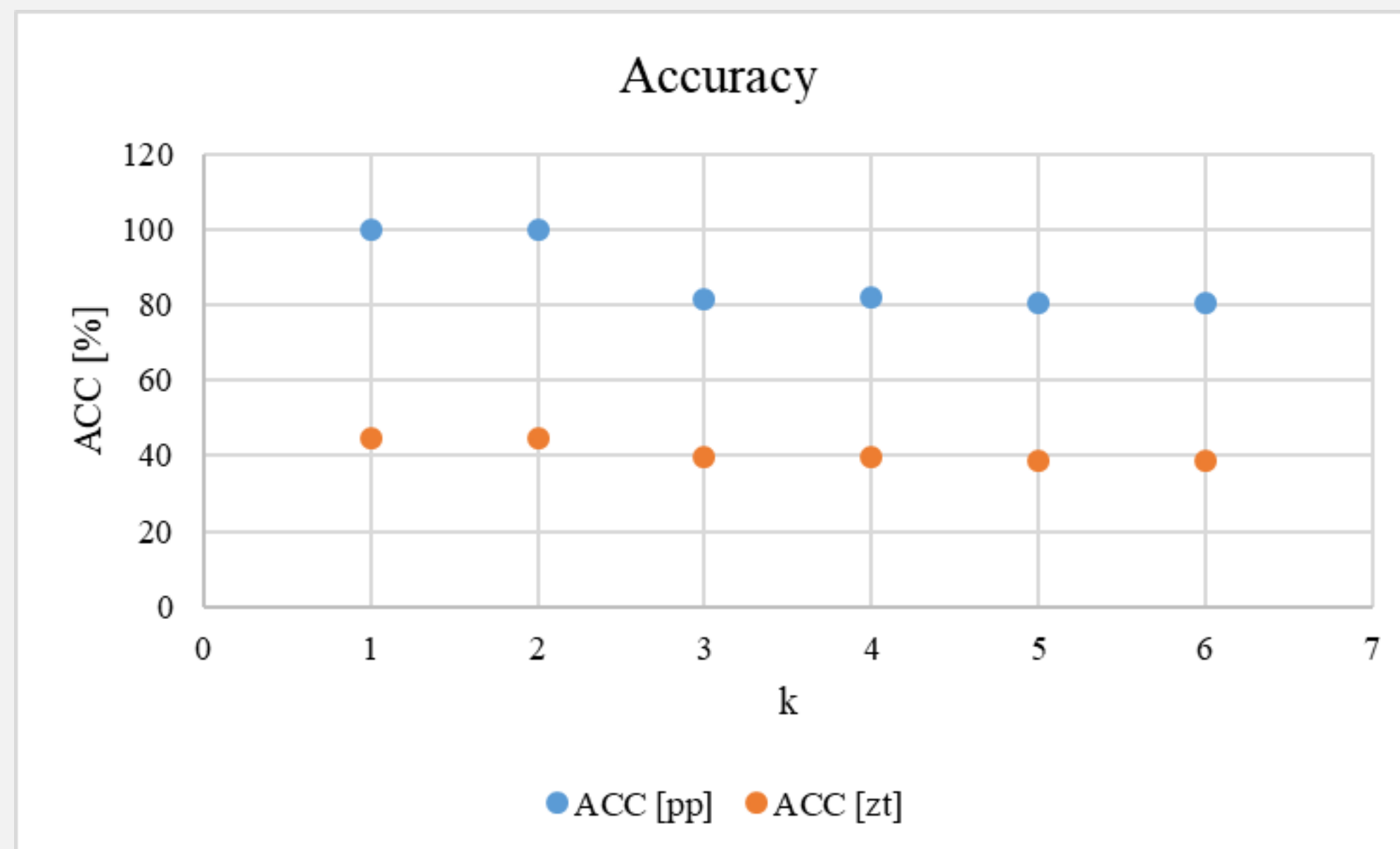
# Klasyfikacja

- Do klasyfikacji wykorzystano 3 metody uczenia pod nadzorem – metoda najbliższych sąsiadów (KNN), LDA, QDA.
- W celu określenia jakości klasyfikatorów skorzystano z macierzy pomyłek, na jej podstawie obliczono dokładność (Accuracy) badanego klasyfikatora.



# KNN (k najbliższych sąsiadów)

- Jest dość prostym sposobem na bezpośrednie określenia przynależności danych obserwacji do poszczególnych klas.
- Wykorzystuje następującą regułę: do danego punktu w przestrzeni przypisz taką klasę, jaką ma większość jego k sąsiadów, przy czym liczba k najbliższych sąsiadów jest z góry narzucona.



# LDA (liniowa analiza dyskryminacyjna)

## QDA (kwadratowa analiza dyskryminacyjna)

- LDA jest używana do znalezienia liniowej kombinacji cech, które najlepiej rozróżniają dwie lub więcej klas obiektów lub zdarzeń. Wynikowe kombinacje są używane jako klasyfikator liniowy.
- **Dokładność:**
  - LDA: 43.5%
  - QDA: 38.6%

# Podsumowanie

- Użyto metod statystycznej eksploracji danych do klasyfikacji odmian win.
- Ocena jakości klasyfikatorów została wykonana na próbie testowej uzyskanej z telefonu innego modelu, niż dane do próby uczącej.
- Dokładność klasyfikatorów jest nieco wyższa od losowego określenia typu dźwigu.

	<b>Dokładność [%]</b>
KNN	44,7
LDA	43,5
QDA	38,6

Rys. 6. Dokładność uzyskanych klasyfikatorów

# Wykorzystane źródła

1. <http://techfresh.pl/jak-dziala-akcelerometr-w-telefonie/>, dostęp: 21.01.2020r.
2. <http://www.if.pw.edu.pl/~julas/LSED/lshed4.html>, dostęp: 21.01.2020r.
3. <http://www.if.pw.edu.pl/~julas/LSED/lshed3.html>, dostęp: 21.01.2020r.

Dziękujemy za uwagę!