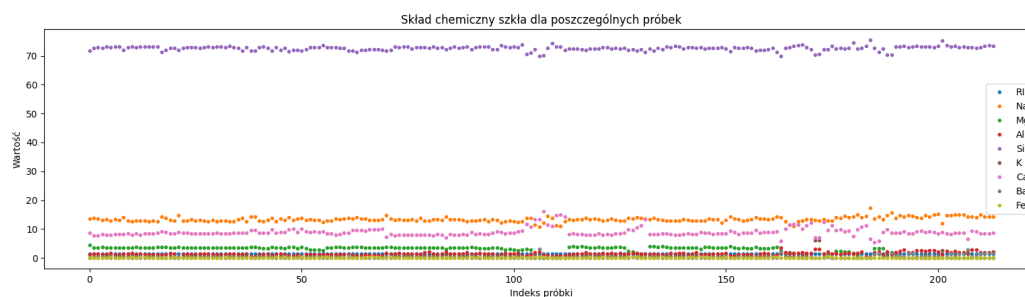


## Zadanie 2 - Maciej Gogulski

### 1. Dla dwóch wybranych zbiorów danych ich wizualizacja przy pomocy wykresów 2D i 3D

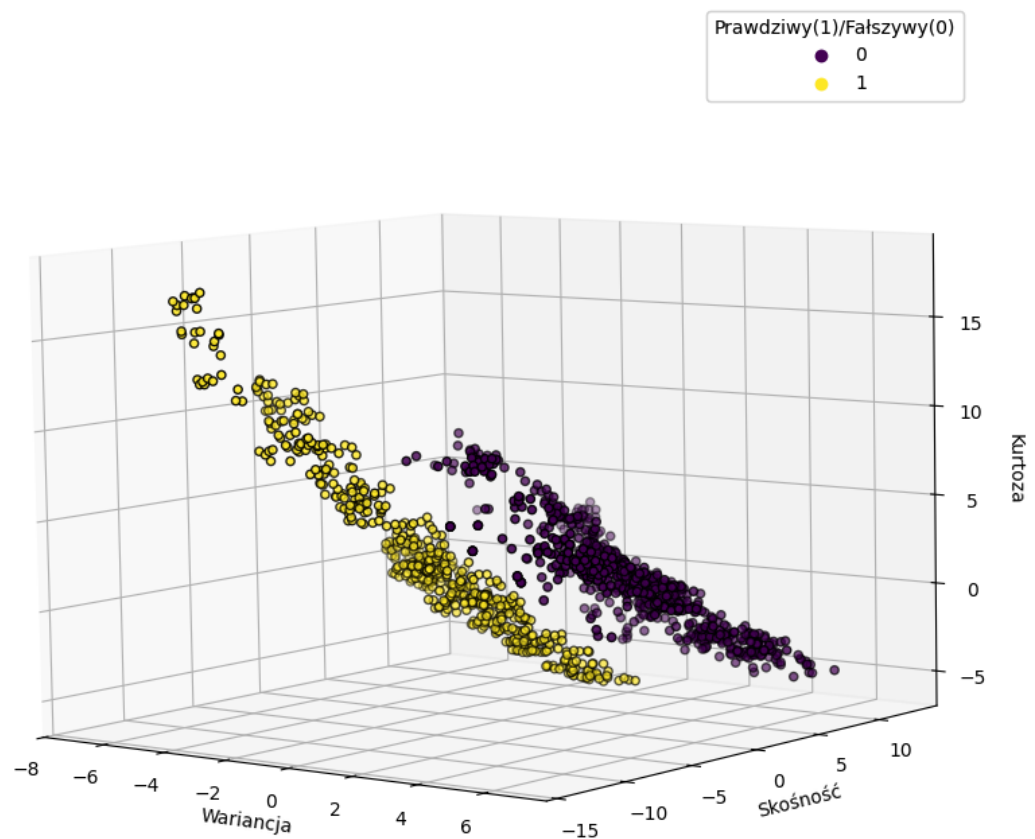
Na potrzeby wizualizacji 2D użyto **Glass Identification** dataset.



W zdecydowanej większości rekordy z datasetu mają bardzo podobne wartości. Istnieją jednak wyjątki, które pokazują pewną zależność cech od siebie. Na wykresie widać wpływ odchyłek od "normy" wartości poszczególnych pierwiastków chemicznych na odchylenia współczynnika załamania światła **RI**.

Na potrzeby wizualizacji 3D użyto **Banknote Authentication** dataset.

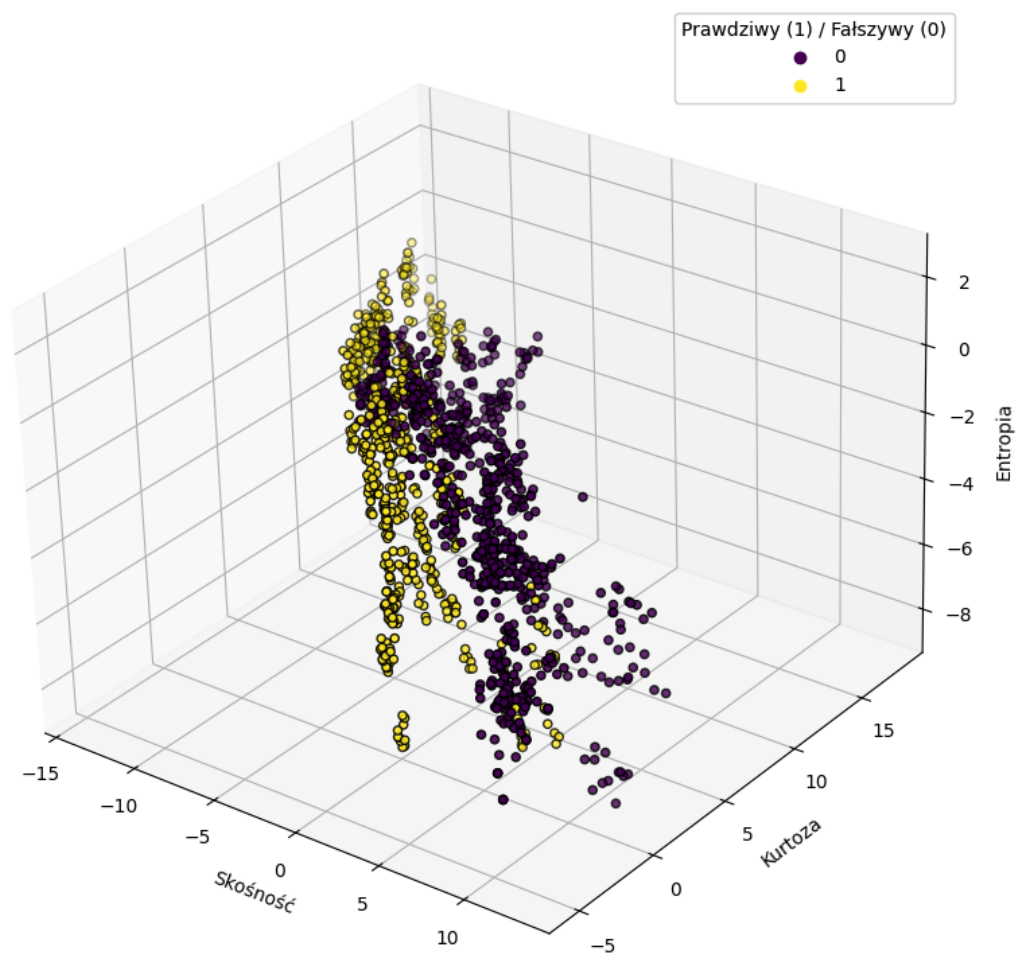
## Wizualizacja 3D: Identyfikacja fałszywych banknotów



Wizualizacja trzech z czterech cech (**wariancja**, **skośność** i **kurtoza**) na trójwymiarowym wykresie oraz pokolorowanie rekordów w zależności od klasyfikacji, uwidoczniło podział obu klas na klastry.

Sprawdzono również co się stanie, gdy dobrany zostanie inny zestaw cech.

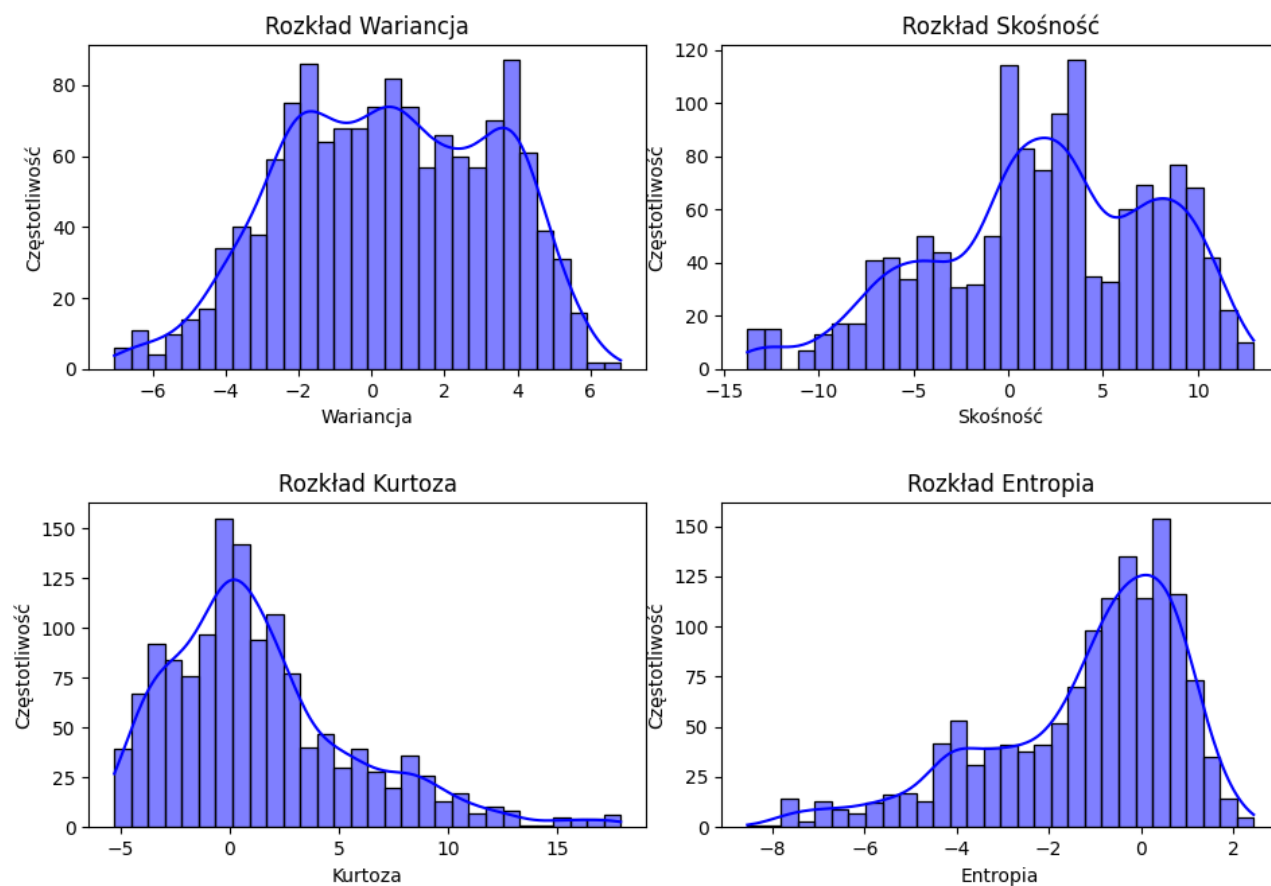
Wizualizacja 3D: Identyfikacja fałszywych banknotów



Przy zestawie **Skośność**, **Kurtoza** i **Entropia** klasteryzacja jest dużo bardziej skomplikowana. Klasy nie są tak wyraźnie odseparowane od siebie jak przy poprzednim zestawie.

## 2. Wizualizacja wybranych atrybutów przy użyciu histogramu.

Pokazano histogramy dla wszystkich czterech cech danych **Banknote Authentication**



Z powyższych histogramów można wywnioskować, że:

- **Wariancja** ma rozkład podobny do normalnego. Oznacza to, że większość banknotów jest podobna do siebie pod względem tej cechy.
- **Skośność** ma dwa wyraźne piki. Oznacza to, że dane mogą pochodzić z dwóch różnych grup.
- **Kurtoza** i **Entropia** mają wyraźny pik w jednej części wykresu. To może wskazywać istnienie nietypowych przypadków.