

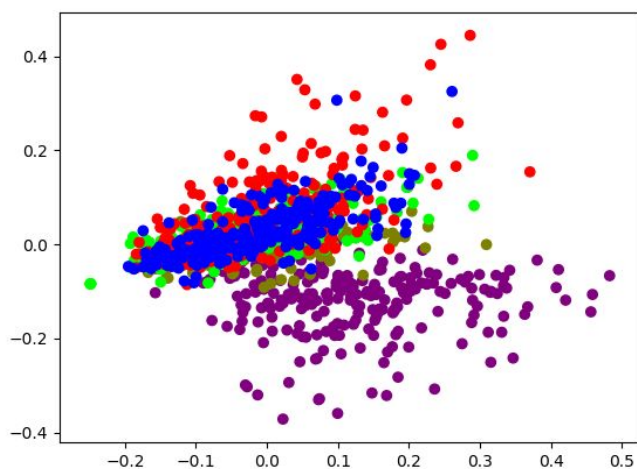
Wizualizacja metodą t-SNE

Józef Jasek

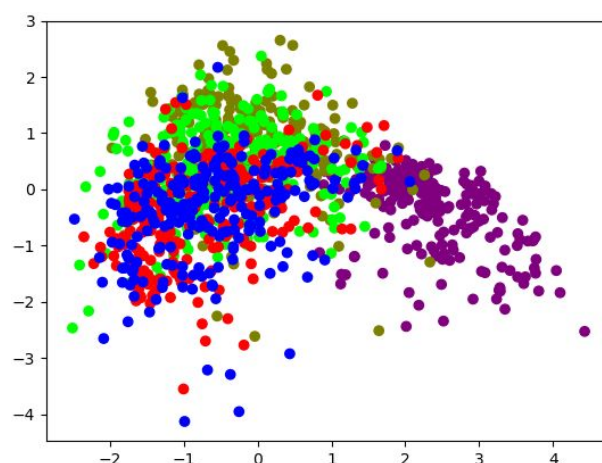
Zadanie 1

Analizowany zbiór to 20 news group. Wybieramy 5 grup i połowę danych z tych zbiorów.

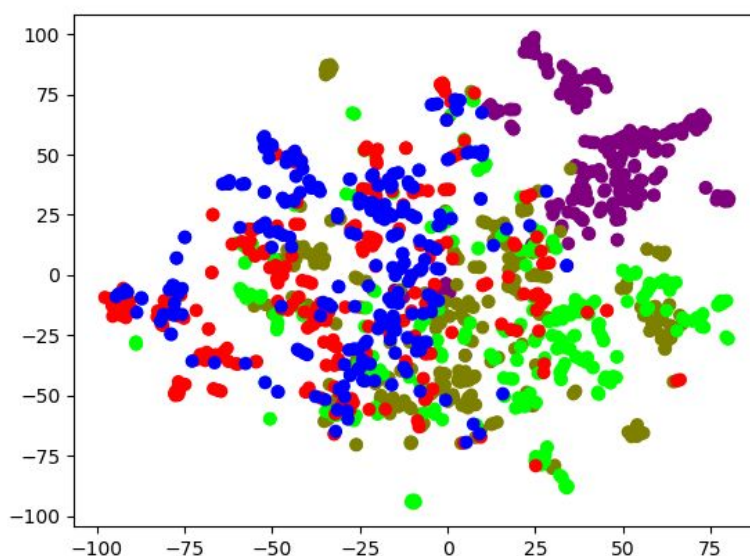
Porównujemy metodę t-SNE do pozostałych posługując się tymi samymi metrykami co w poprzednich sprawozdaniach. Przed zastosowaniem t-SNE na zbiorze użyto PCA.



PCA



ISOMAP



t-SNE (perplexity = 10)

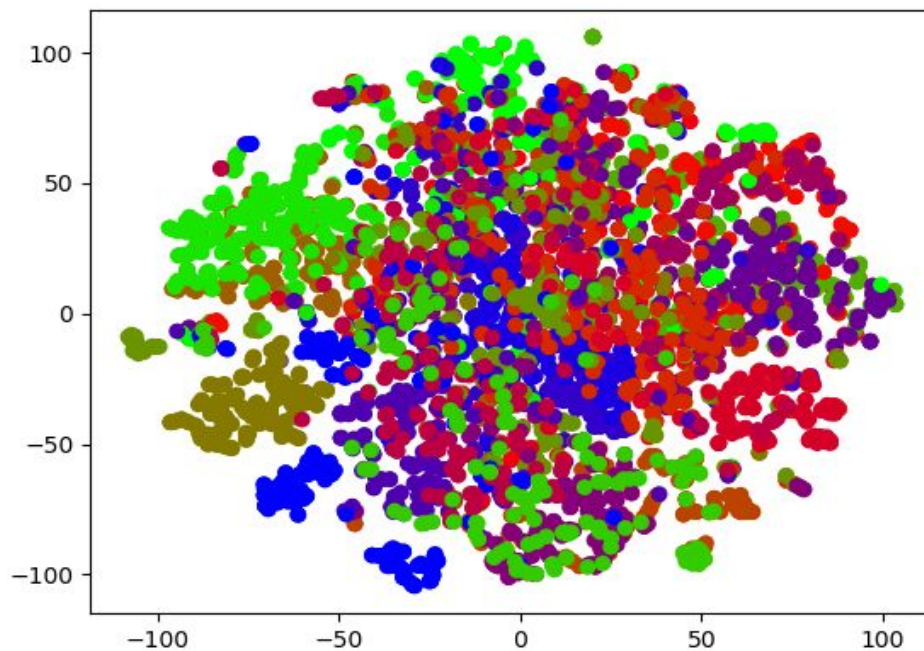
Widzimy, że zbiory są na tyle podobne do siebie, że nie mamy możliwości dobrego wydzielenia grup, jednak t-SNE dużo lepiej sobie w takiej sytuacji radzi z zachowaniem tych samych najbliższych sąsiadów. ($k = 10$)

tsne:	5.283950617283951/10	65%
pca:	1.3657407407407407/10	47%
kpca:	1.3665123456790123/10	47%
mds:	1.158179012345679/10	31%
lle:	1.6481481481481481/10	48%

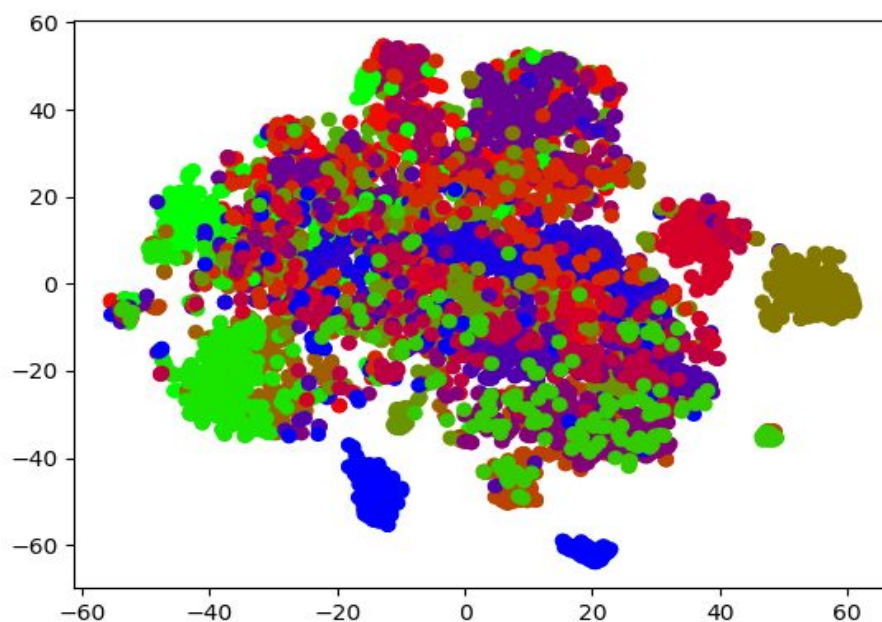
Zadanie 2

W drugiej części zbadamy wpływ perplexity na jakość wizualizacji przy użyciu wszystkich 20 grup ze zbioru 20 news group. Otrzymujemy:

Perplexity 5 :	4.674454828660436/10	51%
Perplexity 20 :	4.963979750778816/10	51%
Perplexity 50 :	4.826518691588785/10	51%
Perplexity 80 :	4.663551401869159/10	50%
Perplexity 120 :	4.426207165109035/10	50%
Perplexity 180 :	4.08547507788162/10	49%



t-SNE (perplexity = 5)



t-SNE (perplexity = 50)