

# Rozpoznawanie cyfr

Etap pierwszy - metoda KNN

# Opis danych

Jest to podzbiór większego zestawu dostępnego w NIST. Cyfry zostały znormalizowane pod względem wielkości i wyśrodkowane na obrazie o stałym rozmiarze (28x28 pikseli).

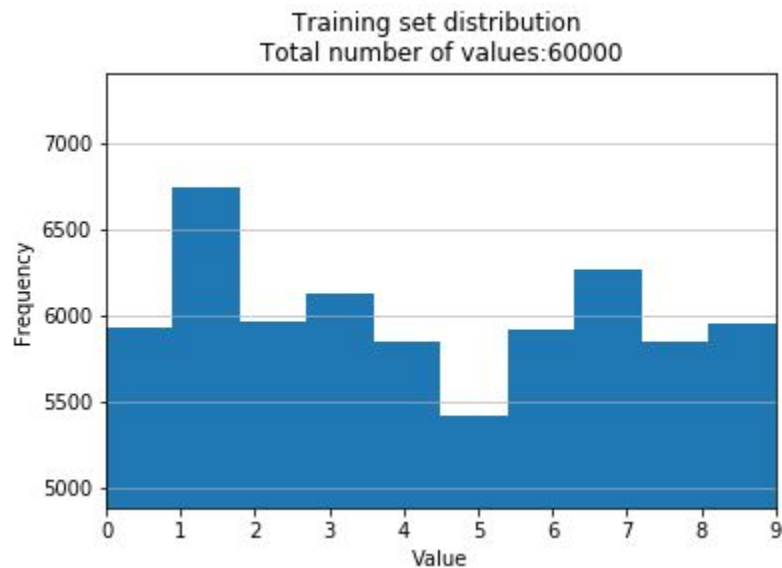
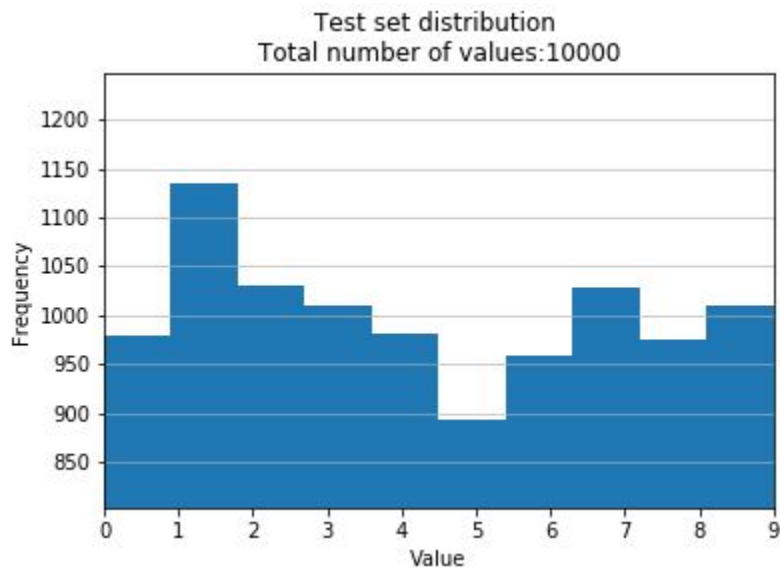
Dane są dostępne w formacie binarnym.

Zbiór treningowy: 60 000

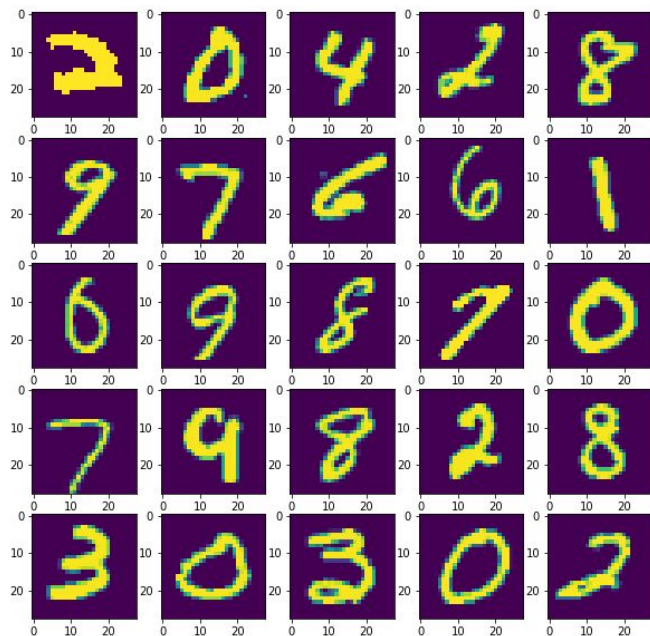
Zbiór testowy: 10 000



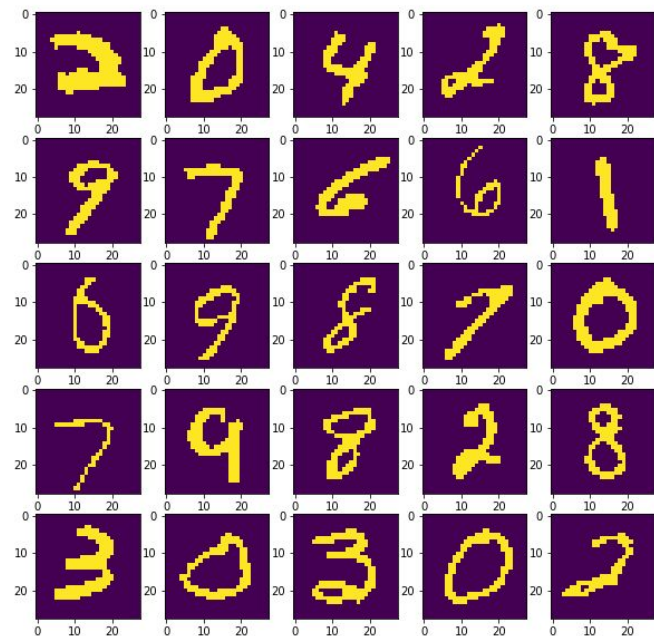
# Rozkład wartości



# Feature extraction - obraz binarny



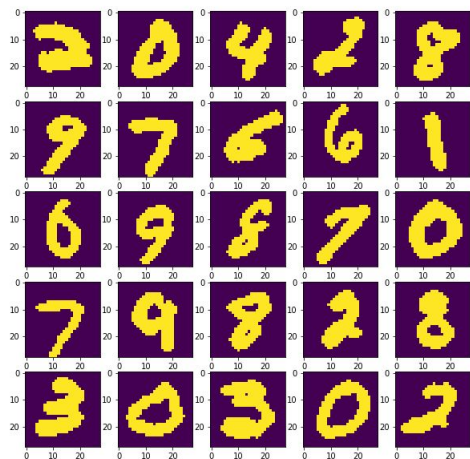
Obraz przed progowaniem



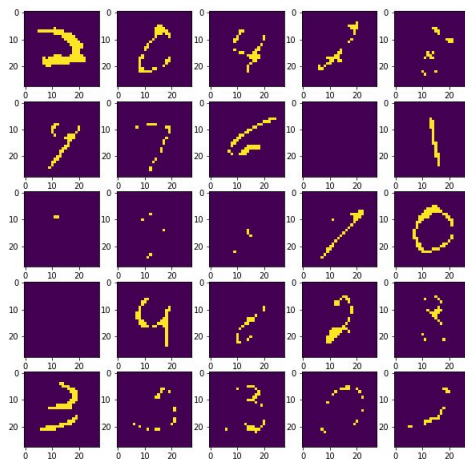
Obraz po progowaniu  $t=150$

# Preprocessing

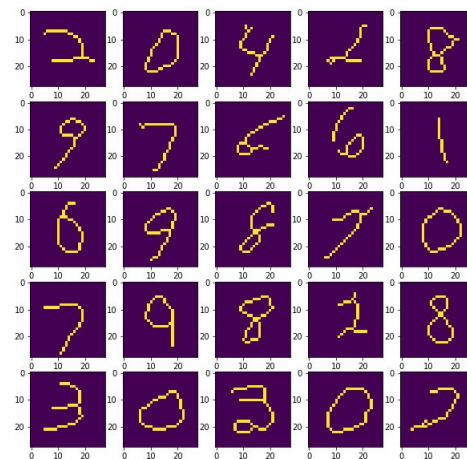
Efekty zastosowaniu różnych filtrów, operacji na obrazie, ścienianie (skeletonization).



Filtr gaussa



Erozja



Ścienianie

**Żadna z nich (również w połączeniach) nie poprawiła wyniku końcowego w istotnym stopniu**

# Wady i zalety modelu KNN

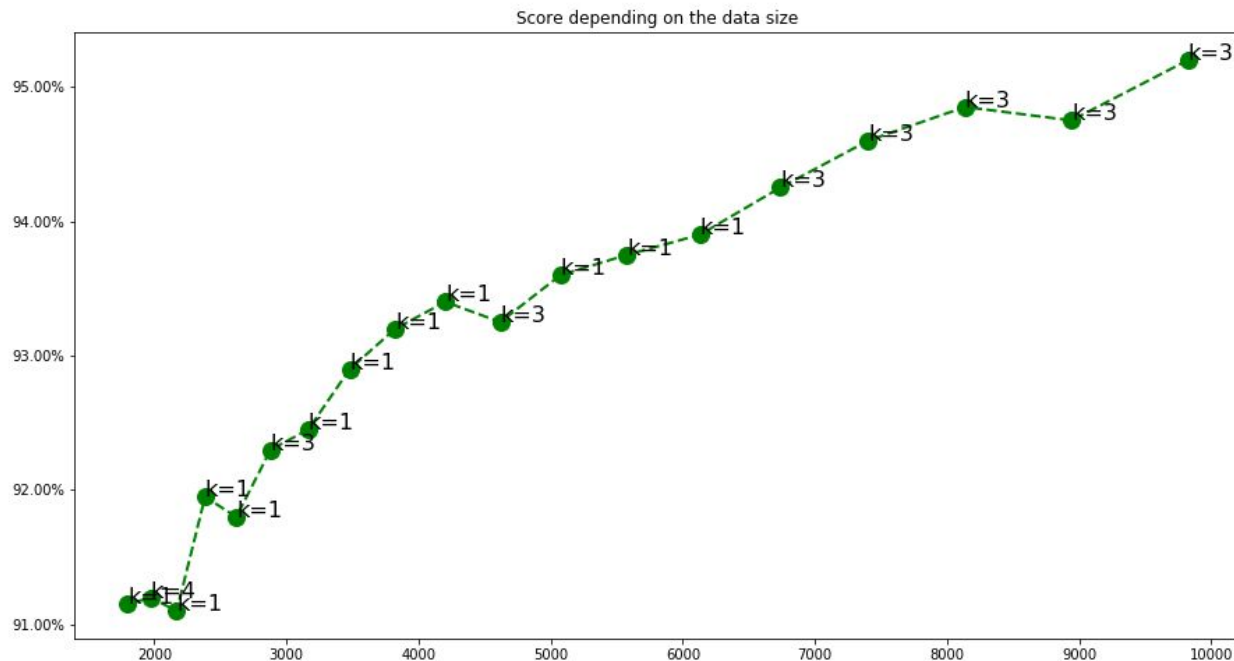
## Plusy

- Prosty algorytm
- Szybkie uczelnie
- Działa dla danych nieliniowych

## Minusy

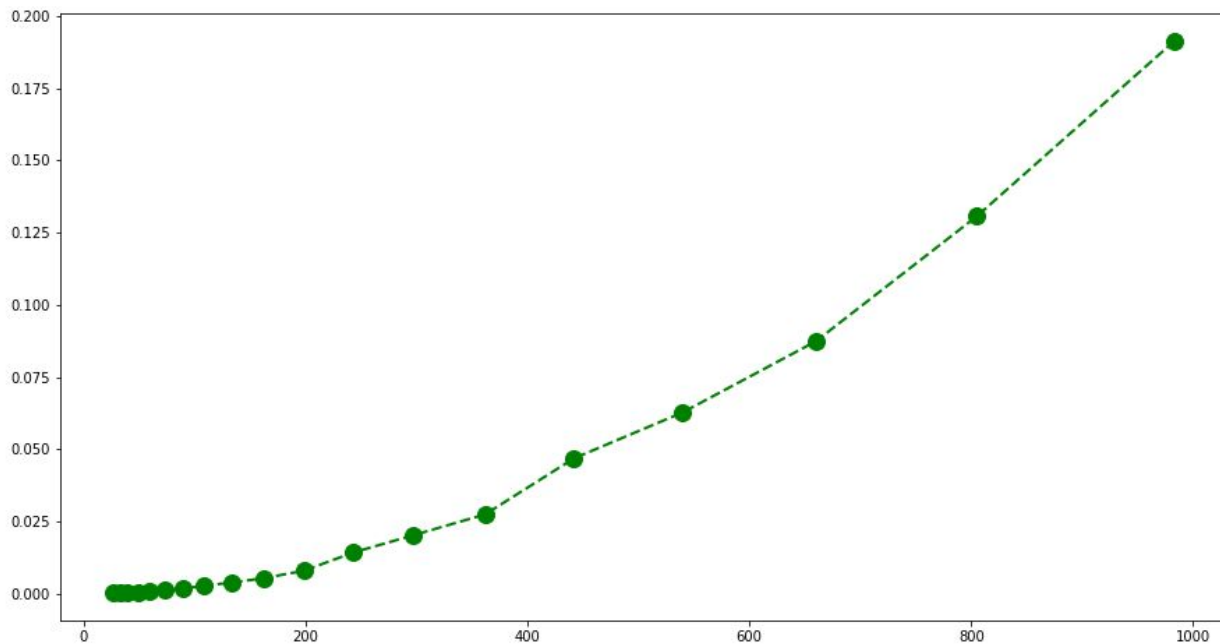
- Kosztowne obliczeniowo
- Wysokie zapotrzebowanie na pamięć
- Wola predykcja

# Dokładność w zależności od zbioru testowego





# Predykcje KNN są bardzo wolne



# Wyniki

Classification report:

	precision	recall	f1-score	support
0	0.97	0.99	0.98	176
1	0.92	1.00	0.96	236
2	0.95	0.96	0.96	211
3	0.93	0.93	0.93	194
4	0.95	0.94	0.95	198
5	0.96	0.95	0.96	198
6	0.98	0.98	0.98	191
7	0.93	0.92	0.93	181
8	0.98	0.91	0.94	200
9	0.94	0.92	0.93	215
accuracy			0.95	2000
macro avg	0.95	0.95	0.95	2000
weighted avg	0.95	0.95	0.95	2000

Otrzymaliśmy klasyfikator o  
predykcyjności 95% na zbiorze testowym

