

Klasyfikacja kształtu dłoni w grze “Kamień, Papier, Nożyce”

Przedmiot: Podstawy reprezentacji i analizy danych
Autorzy: Martyna Jakubowska, Hubert Nakielski, Kacper Seredyn

Poruszany problem

Celem projektu jest klasyfikacja gestów “Kamień, Papier, Nożyce” na podstawie zdjęć dłoni.

Zbiór danych

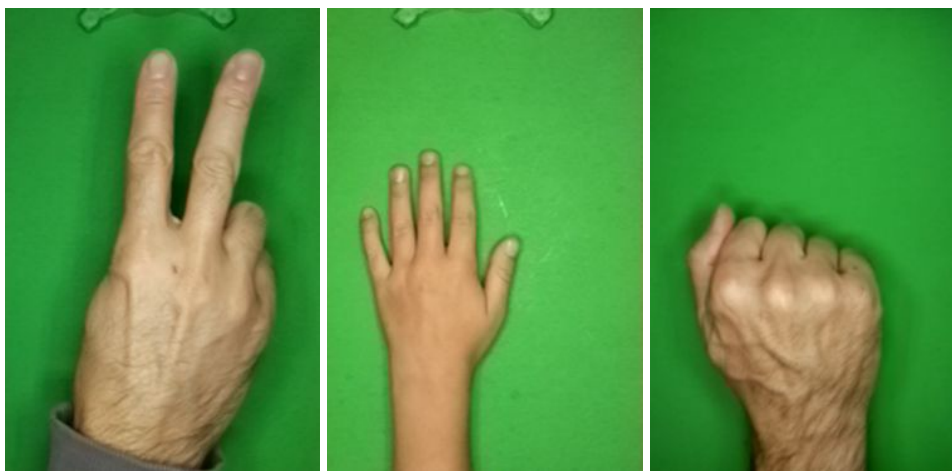
Zestaw zawiera łącznie 2188 obrazów odpowiadających gestom dłoni z gry “Kamień, Papier, Nożyce”:

- “Kamień” (726 obrazów),
- “Papier” (710 obrazów),
- “Nożyce” (752 obrazy).

Wszystkie zdjęcia są robione na zielonym tle ze względnie spójnym oświetleniem i balansem bieli. Wszystkie obrazy są obrazami RGB o szerokości 300 pikseli i wysokości 200 pikseli w formacie .png.

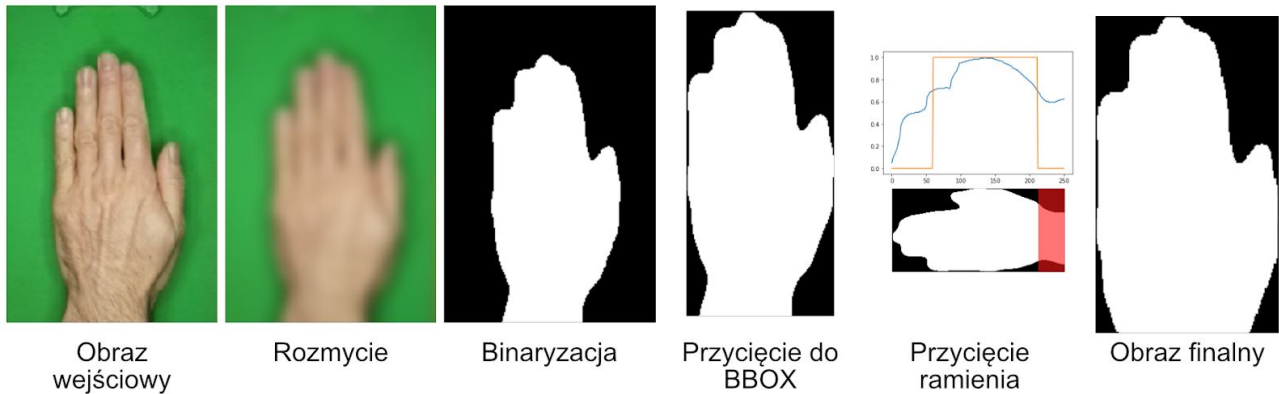
Obrazy są podzielone na trzy podfoldery o nazwach 'rock' (kamień), 'paper' (papier) oraz 'scissors' (nożyce) zgodnie z ich klasą.

Przykładowe zdjęcia ze zbioru:



Przygotowanie danych

Obróbka obrazów



Binaryzacja

Przed dokonaniem izolacji dłoni od tła, rozmywamy obraz przy użyciu filtra Gaussa, aby zapobiec wysokoczęstotliwościowym artefaktom w zbinaryzowanych danych.

Początkowo, do binaryzacji została użyta metoda progowa korzystająca z danych koloru przekonwertowanych na HSV, aby uwzględnić nie tylko dłonie, ale także i rękawy znajdujące się na niektórych zdjęciach. W trakcie testowania okazało się jednak, że zastosowanie metody opierającej się o intensywność kanału czerwonego RGB i zignorowanie rękawów spowodowało zwiększenie efektywności klasyfikacji (dla drzewa decyzyjnego $n = 5$, około o 3% wyższy sklearn accuracy score).

Przycinanie

Do dalszego przetwarzania musimy przyciąć obraz tak, aby pozostała w nim tylko dłoń.

Dokonyjemy tego w dwóch etapach:

1. Przycięcie do minimalnego prostokąta zawierającego obiekt w obrazie - stosujemy w tym celu **bounding box** obiektu.
2. Wycięcie pozostałego fragmentu ramienia - dla każdego wiersza obrazu liczona jest średnia wartość pikseli tego wiersza. Następnie, idąc od dołu, szukany jest pierwszy wiersz, którego średnia wartość przekracza pewną stałą (przyjętą w projekcie na 0,6). Wszystkie wiersze poniżej znalezionej zostają odrzucone.



Ekstrakcja cech

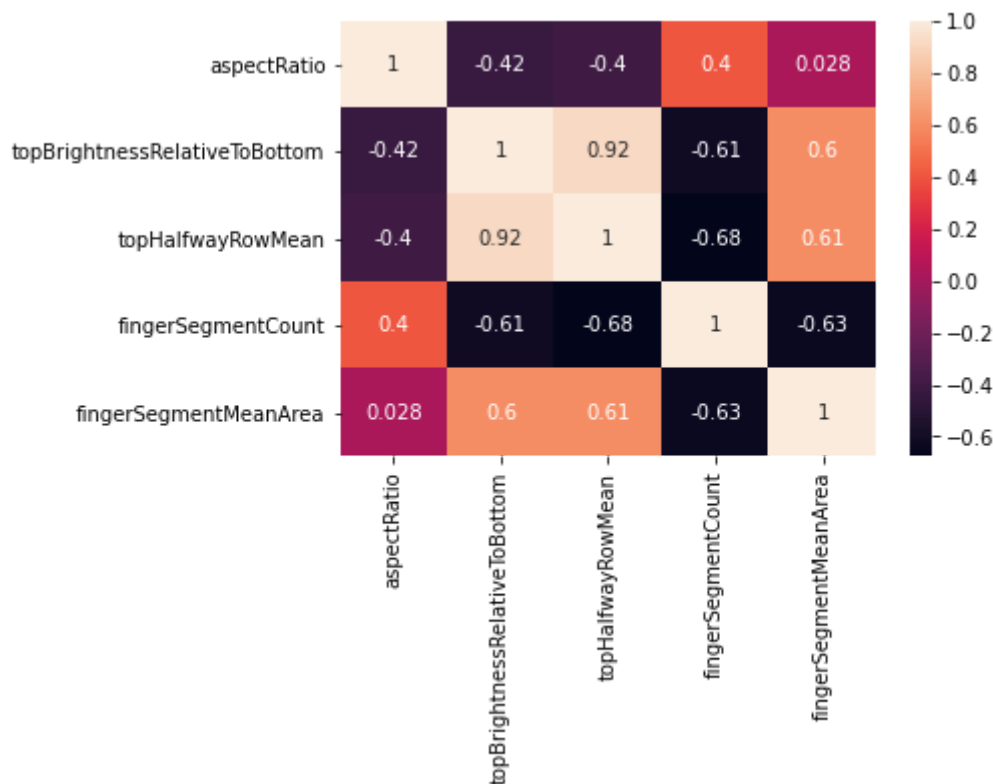
Z przygotowanego obrazu binarnego dokonujemy ekstrakcji następujących cech:

- **Współczynnik proporcji** (aspectRatio) - iloraz szerokości do wysokości obrazu.
Hipoteza: Współczynnik proporcji obrazu pozwoli klasyfikatorowi stwierdzić, czy palce są wyciągnięte, czy też nie, co pozwoli odróżnić ułożenie dłoni w "Kamień" od pozostałych gestów.
(diagram)
- **Względna "jasność" górnej części obrazu** (topBrightnessRelativeToBottom) - obraz przecinany jest linią poziomą na dwie równe części (górną i dolną). Następnie liczony jest iloraz średniej jasności górnej połowy zdjęcia przez średnią jasność dolnego fragmentu.
Hipoteza: Dzięki obliczonej względnej jasności górnej części obrazu klasyfikator będzie w stanie odróżnić rozstawienie "Papier" od ułożenia "Nożyce".
- **Średnia wartość wiersza w połowie górnej części obrazu** (topHalfwayRowMean) - średnia wartość pikseli wiersza znajdującego się w połowie górnej połowy obrazu.
Hipoteza: Powyższa cecha wspomogę odróżnienie rozstawienie "Papier" od "Nożyce".
- **Liczba segmentów górnej ćwierci obrazu oraz ich średnie pole** (fingerSegmentCount, fingerSegmentMeanArea) - górna połowa obrazu jest ponownie podzielona poziomo na pół, w górnej ćwierci liczone są segmenty dłoni oraz ich średnie pole (ilość pikseli w każdym segmencie).
Hipoteza: Podobnie jak w poprzednim punkcie, powyższe cechy wspomogą odróżnianie papieru od nożyc.

Analiza przydatności atrybutów

W celu sprawdzenia, które z atrybutów będą przydatne w klasyfikacji, utworzone zostały następujące wykresy:

- Wykres korelacji



- Macierz wykresów punktowych



Wnioski:

- Atrybuty **topBrightnessRelativeToBottom** oraz **topHalfwayRowMean** są ze sobą dość ściśle skorelowane.
- Atrybut **fingerSegmentCount** ze względu na swoją dyskretną naturę oraz wysoką podatność na zmiany związane z małymi zmianami pozycji palców nie jest przydatny w analizie.
- Atrybut **fingerSegmentMeanArea** dobrze radzi sobie z wyróżnieniem ułożenia “nożyce”, lecz ułożenia “papier” oraz “kamień” nakładają się na siebie. Ponadto, w części przypadków, ułożenie “nożyce” pokrywa się z “papier”, prawdopodobnie z powodu rozłożonych palców dla pozy papier.

Ostatecznie, cecha **fingerSegmentCount** została wyeliminowana z dalszej analizy.

Klasyfikacja

Uczenie nadzorowane

Dla danego zbioru danych, najlepszym sposobem klasyfikacji jest uczenie nadzorowane. Dla porównania użyte zostaną następujące klasyfikatory:

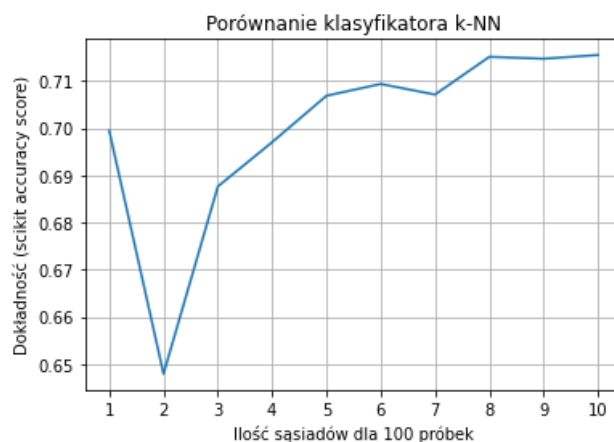
- k-Najbliższych sąsiadów (k-NN)
- Najbliższego prototypu
- Naiwny klasyfikator Bayesa
- Drzewo decyzyjne (DT)

Dobór parametrów klasyfikatorów

Przed porównaniem klasyfikatorów między sobą, najpierw zostały dobrane optymalne parametry każdego z klasyfikatorów.

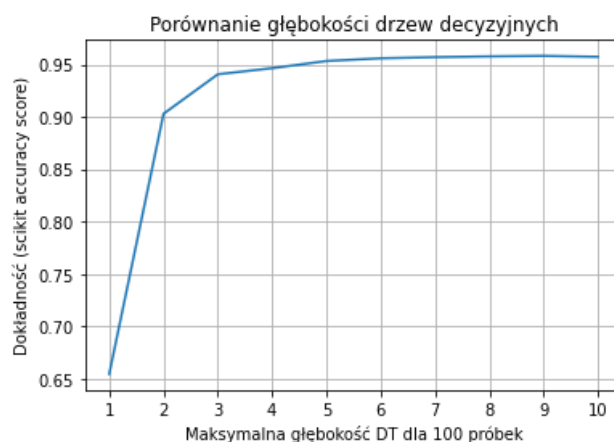
k-Najbliższych sąsiadów

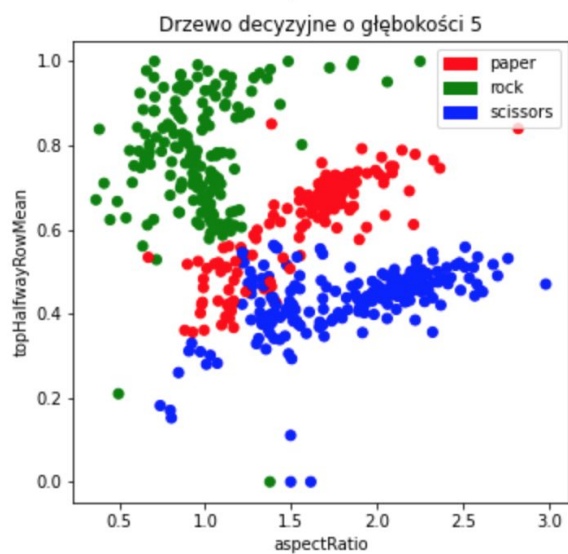
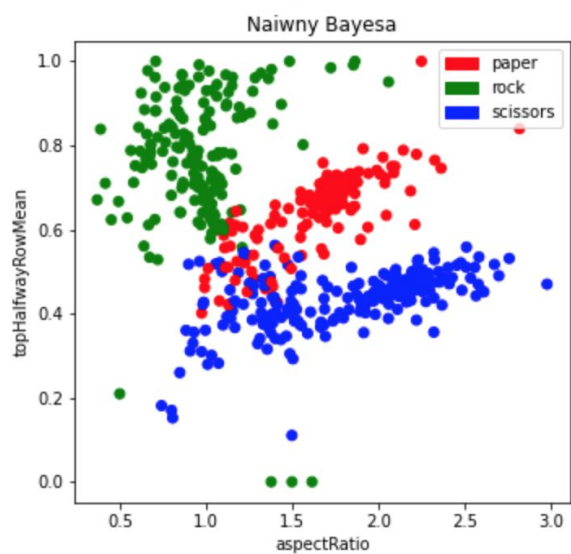
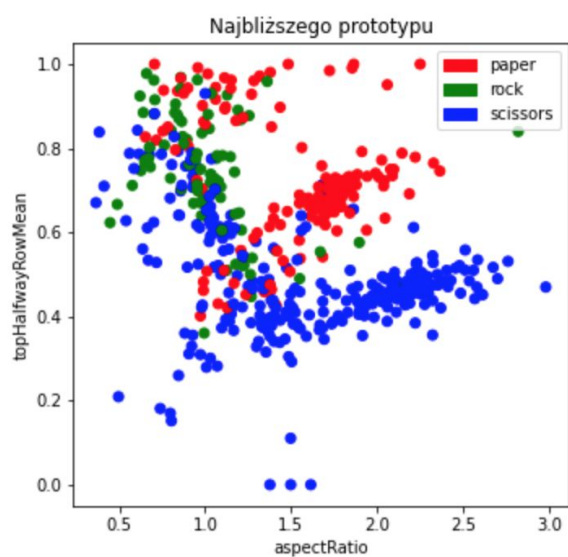
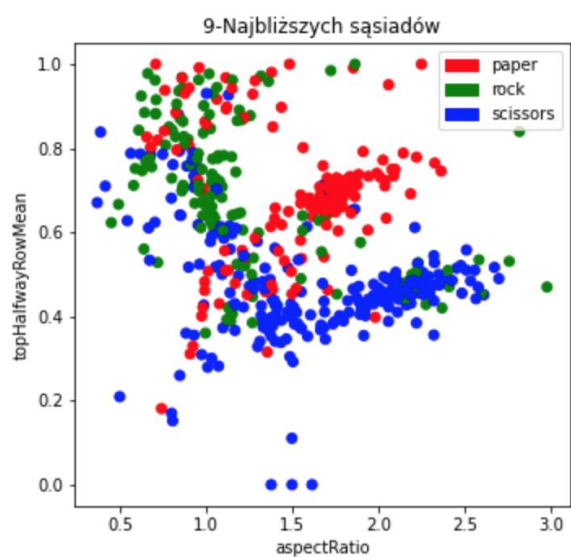
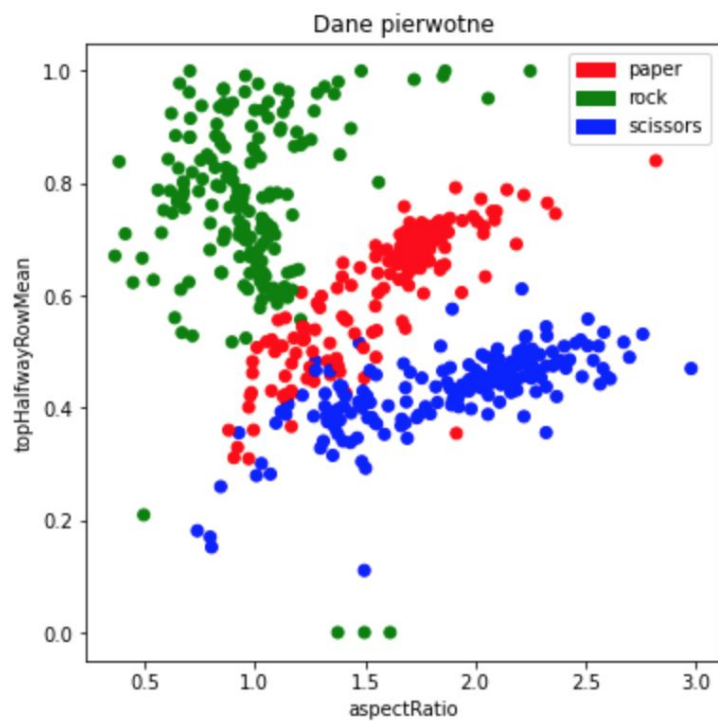
Sprawdzona została trafność klasyfikacji klasyfikatora k-NN dla k od 1 do 10. Trafność klasyfikatora k-NN była mocno uzależniona od podziału danych, więc kształt wykresu zmieniał się znacznie między uruchomieniami. Ostatecznie wartość k została wybrana na **k = 9**.



Drzewo decyzyjne

Sprawdzona została trafność klasyfikacji dla głębokości od 1 do 10 drzewa decyzyjnego. Na podstawie tych danych, optymalna głębokość została ustalona jako **n = 5**, ponieważ dla tego n trafność klasyfikatora utrzymywała się lekko powyżej 95% przy wielu uruchomieniach.





Weryfikacja i porównanie

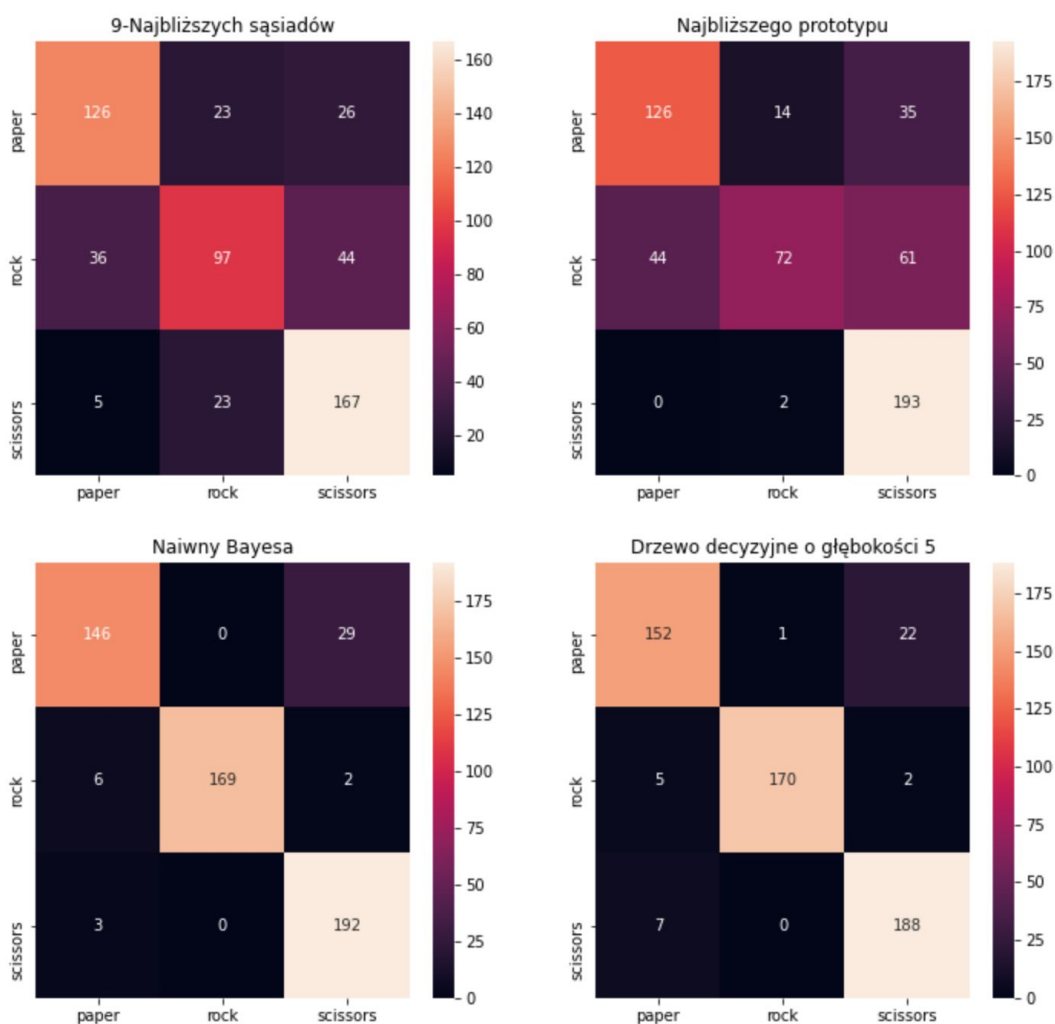
Dla każdego z klasyfikatorów został obliczony accuracy score (uśredniony z 100 losowych podziałów zbioru danych):

Klasyfikator	% poprawnie sklasyfikowanych danych
9-Najbliższych sąsiadów	70.97806215722119
Najbliższego prototypu	69.34917733089581
Naiwny Bayesa	91.61060329067642
Drzewo decyzyjne n = 5	96.14625228519192

Najlepszym wynikiem wykazało się **drzewo decyzyjne o maksymalnej głębokości = 5**.

Dla każdego klasyfikatora została wykonana macierz pomyłek z jednego z losowych podziałów danych. Z niej wynikają następujące wnioski:

- Wszystkie klasyfikatory najlepiej poradziły sobie z gestem “nożyce”.
- Klasyfikatory najbliższych sąsiadów/prototypu lepiej poradziły sobie z gestem “papier” niż “kamień”.
- Klasyfikatory Bayesa oraz DT lepiej poradziły sobie z gestem “kamień” niż “papier”.



Wyniki

W trakcie procesu analizy eksploracyjnej udało się wyeliminować jedną z cech, która mogłaby negatywnie wpłynąć na proces uczenia. Poprzez porównanie różnych klasyfikatorów i ich parametrów, otrzymano pojedynczy klasyfikator (drzewo decyzyjne $n = 5$), który osiągał trafność wynoszącą średnio 95%.

