

Autor: Szymon Tokarz

Data: 19.11.2024 r. Godz. 8.00

Ćwiczenie: Uczenie maszynowe

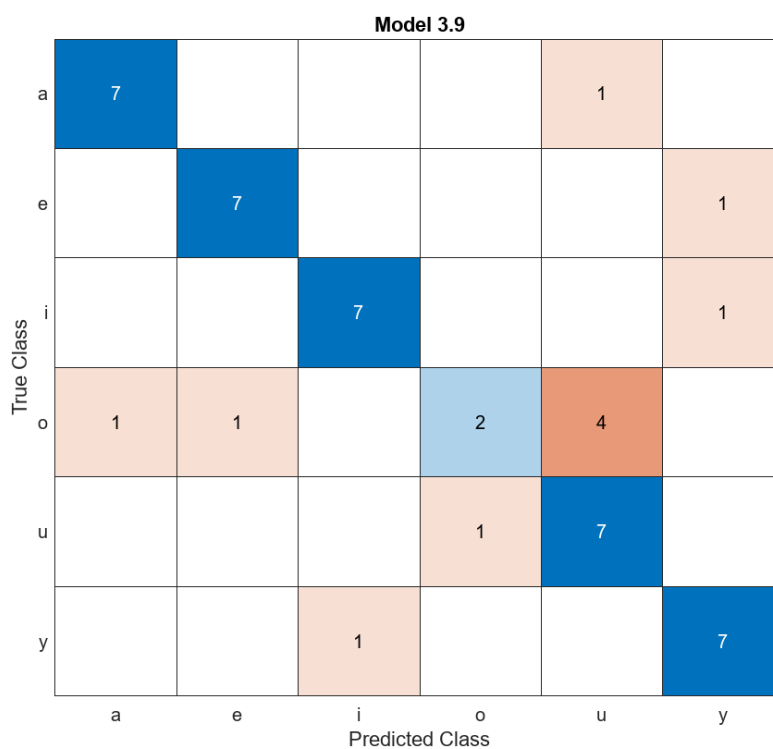
## Rezultaty

### Część 1

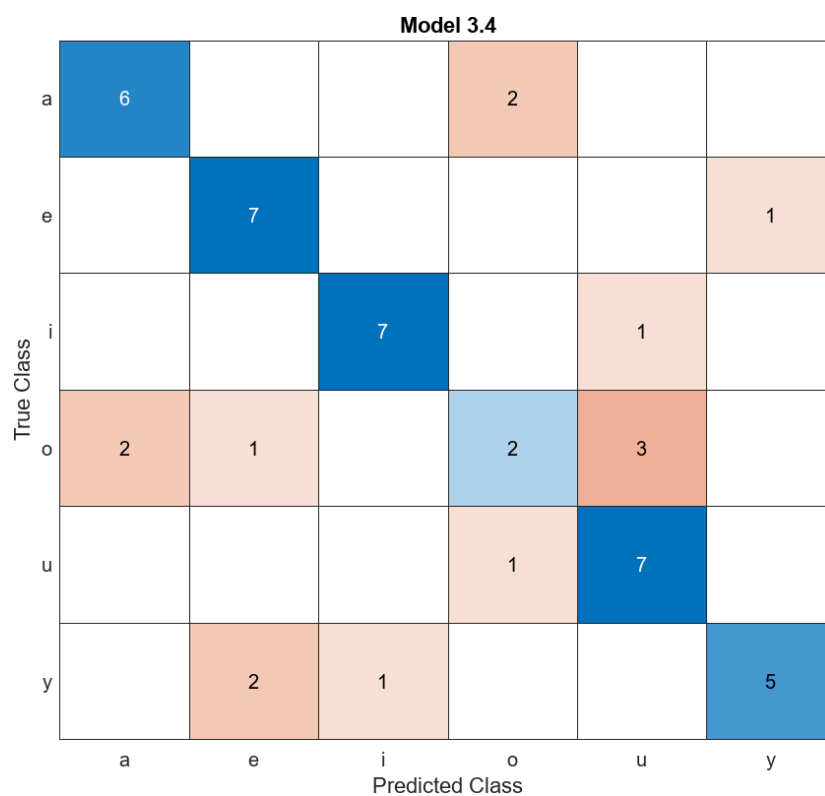
Uzupełniony kod:

### Część 2

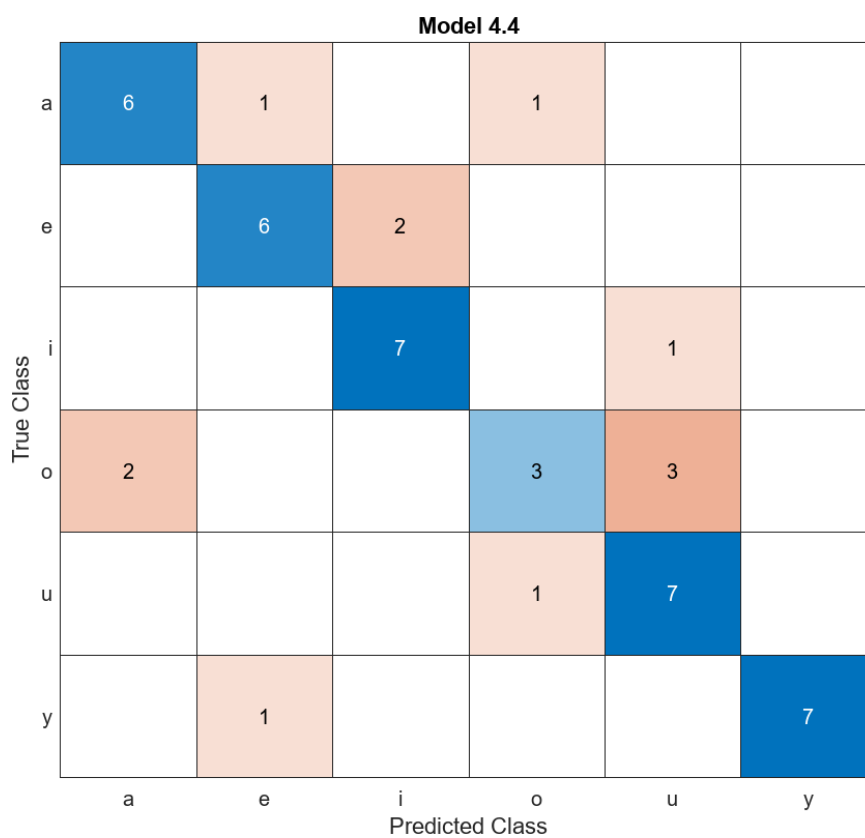
Należało zamieścić macierze pomyłek dla najlepszych dokładności klasyfikacji bez włączonej opcji PCA.



Rys.1 Macierz pomyłek dla dokładności 77.1%



Rys. 2 Macierz pomyłek dla dokładności 70.8%



Rys. 3 Macierz pomyłek dla dokładności 75% z PCA

**Model 2.10**

	a	e	i	o	u	y
a	2	1			1	
e	1	1				2
i			2			2
o				3	1	
u				2	1	1
y		2	1			1
	a	e	i	o	u	y

Predicted Class

Rys. 4 Macierz pomyłek dla dokładności 41.7% dla formantsTableTest

## Analiza i wnioski

Liczebność dla wszystkich samogłosek wynosi 12. Przy podziale 70% i 30% wszystkie samogłoski mają 8 przykładów w zbiorze treningowym i 4 w zbiorze walidacyjnym. Przy podziale walidacyjnym niektóre w zbiorze treningowym mają 8 albo 9 przykładów, lecz jest to wyrównane przez liczbę przykładów w zbiorze walidacyjnym. Ważne powodem, dla którego w uczeniu maszynowym klasy powinny mieć podobną liczebność jest chęć uniknięcia zbytniego skupienia się modelu na elementach neutralnych.

Z macierzy można zauważyć, że najczęściej mylonymi literą jest „o”, która jest mylona z „u” i „a”. Jest to spowodowane podobnymi kształtami tych znaków.

Użycie PCA nie poprawiło dokładności klasyfikacji. Najlepszym klasyfikatorem jest model 3.9 z dokładnością 77.1% nie używający PCA.

Dla klasyfikatora wygenerowanego dla danych treningowych wartości  $\text{validationAccuracy} = 0.7917$  i  $\text{TrainAccuracy} = 1$  różnią się. Jest to spowodowane tym, że zbiór walidacyjny jest inny niż treningowy, przez co klasyfikacja może się różnić.

## Pytania

**Czym różni się prosta walidacja (holdout validation) od ręcznego podziału na zbiór uczący i testowy?**

**Wyjaśnij na czym polega walidacja krzyżowa (k-fold cross-validation) i jaki jest jej cel. Jakie są inne sposoby podziału zbioru danych na zbiór uczący i testowy? Jaką walidację powinno się stosować w przypadku gdy zbiór danych jest niewielki?**

**Wyjaśnij na czym polega specyficzność oraz precyzja klasyfikatora.**

Specyficzność to miara wskazująca w jakim procencie klasa faktycznie negatywna została pokryta przewidywaniem negatywnym.

Precyzja miara wskazująca z jaką pewnością można ufać wskazaniom klasyfikatora.

**Jak jest rola PCA (ang. Principal Component Analysis) w klasyfikacji?**

Usunięcie wielowymiarowości przy zachowaniu najważniejszych cech.

**Dlaczego stosujemy 3 zbiory: uczący, walidacyjny oraz testowy?**

W celu uniknięcia nadmiernego dopasowania do zbioru uczącego.