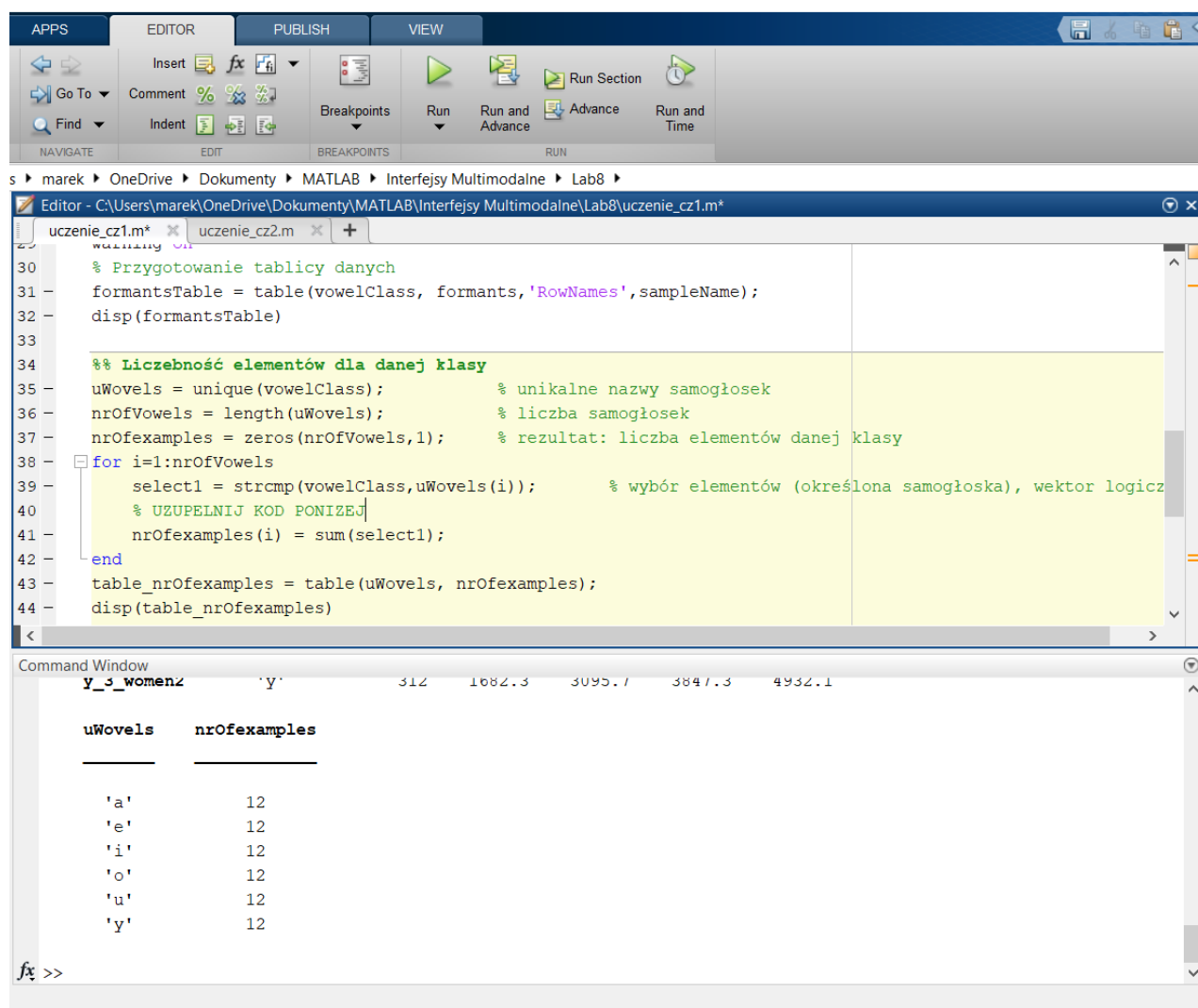


Data: 23.05.2020

Imię i nazwisko: Marek Matys

# Rezultaty

## *Część I – Import i przygotowanie danych*



The screenshot displays the MATLAB environment. The Editor window shows a script named `uczenie_cz1.m` with the following code:

```
30 % Przygotowanie tablicy danych
31 formantsTable = table(vowelClass, formants, 'RowNames', sampleName);
32 disp(formantsTable)
33
34 %% Liczebność elementów dla danej klasy
35 uVowels = unique(vowelClass); % unikalne nazwy samogłosek
36 nrOfVowels = length(uVowels); % liczba samogłosek
37 nrOfexamples = zeros(nrOfVowels,1); % rezultat: liczba elementów danej klasy
38 for i=1:nrOfVowels
39     select1 = strcmp(vowelClass,uVowels(i)); % wybór elementów (określona samogłoska), wektor logiczny
40     % UZUPEŁNIJ KOD PONIZEJ
41     nrOfexamples(i) = sum(select1);
42 end
43 table_nrOfexamples = table(uVowels, nrOfexamples);
44 disp(table_nrOfexamples)
```

The Command Window shows the output of the script. It first displays the variable `y_3_women2` with its dimensions and values. Then, it displays the table `table_nrOfexamples` with the following data:

uVowels	nrOfexamples
'a'	12
'e'	12
'i'	12
'o'	12
'u'	12
'y'	12

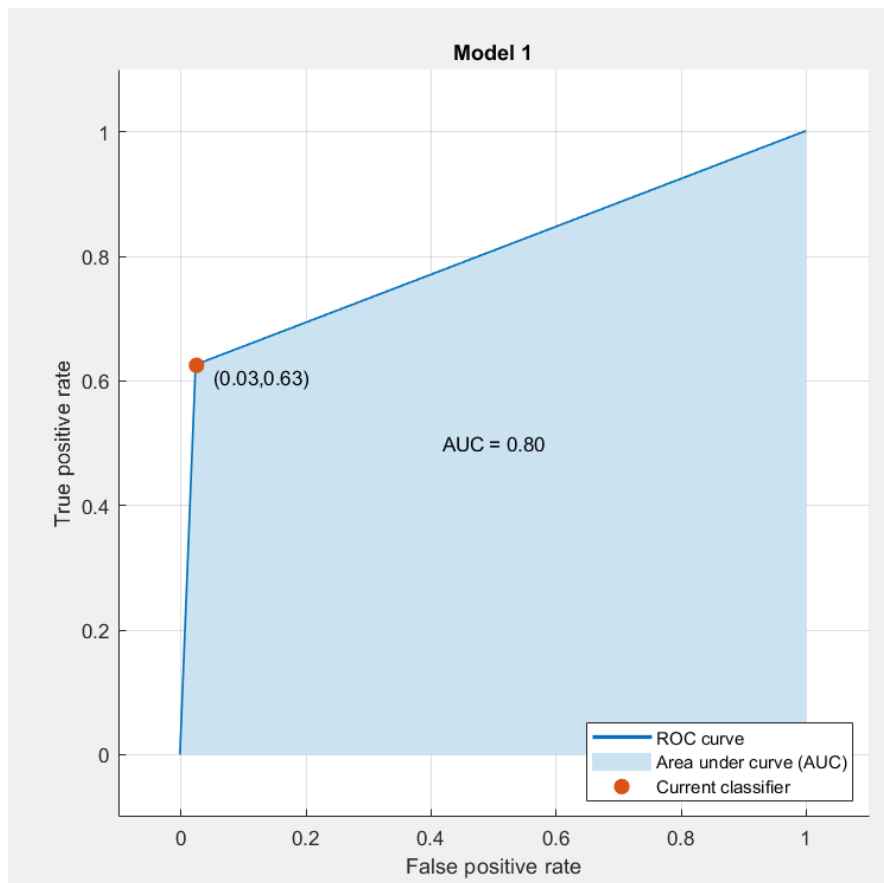
## Część II – budowa modelu (uczenie klasyfikatorów)

▼ History		
1 ☆ KNN	Accuracy: <b>72.9%</b>	
Last change: Fine KNN		5/5 features
2 ☆ SVM	Accuracy: 64.6%	
Last change: Quadratic SVM		5/5 features
3 ☆ Tree	Accuracy: 64.6%	
Last change: Fine Tree		5/5 features

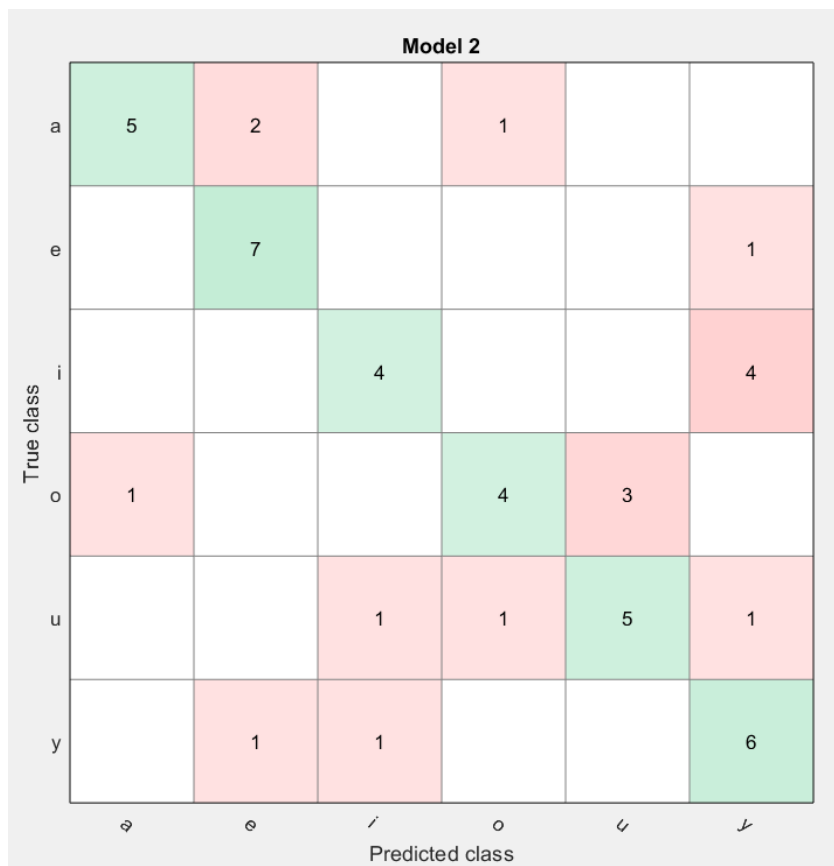
Najlepszy wynik treningu uzyskał klasyfikator KNN

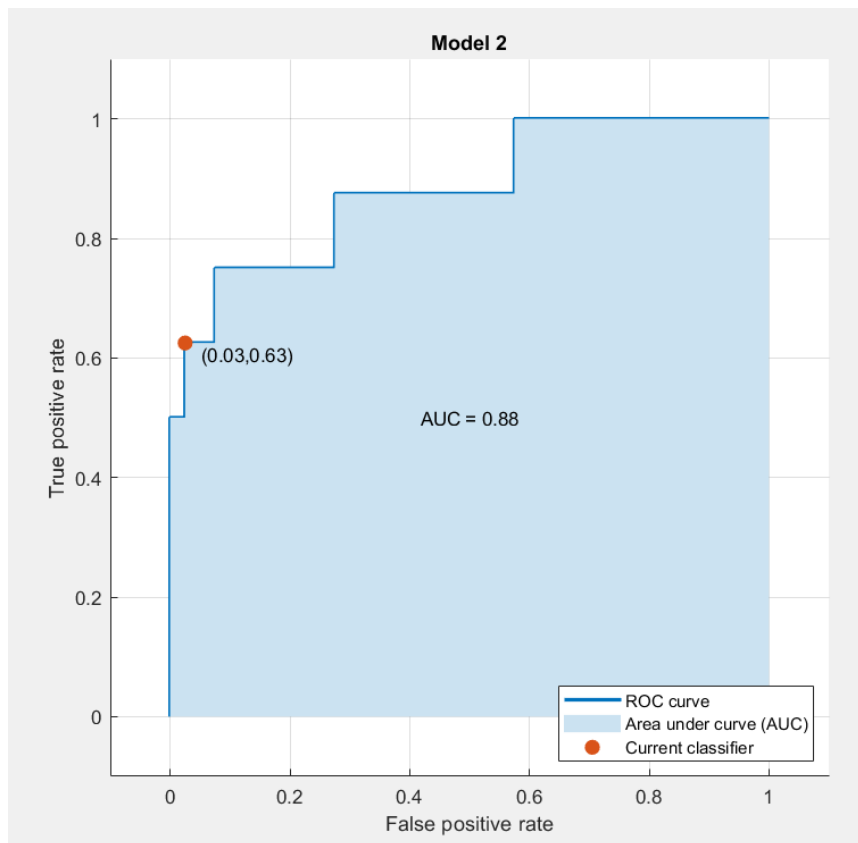
KNN

		Model 1					
True class	a	5			3		
	e		7				1
	i			6			2
	o	1	2		4	1	
	u				1	6	1
	y		1				7
		Predicted class					
		a	e	i	o	u	y

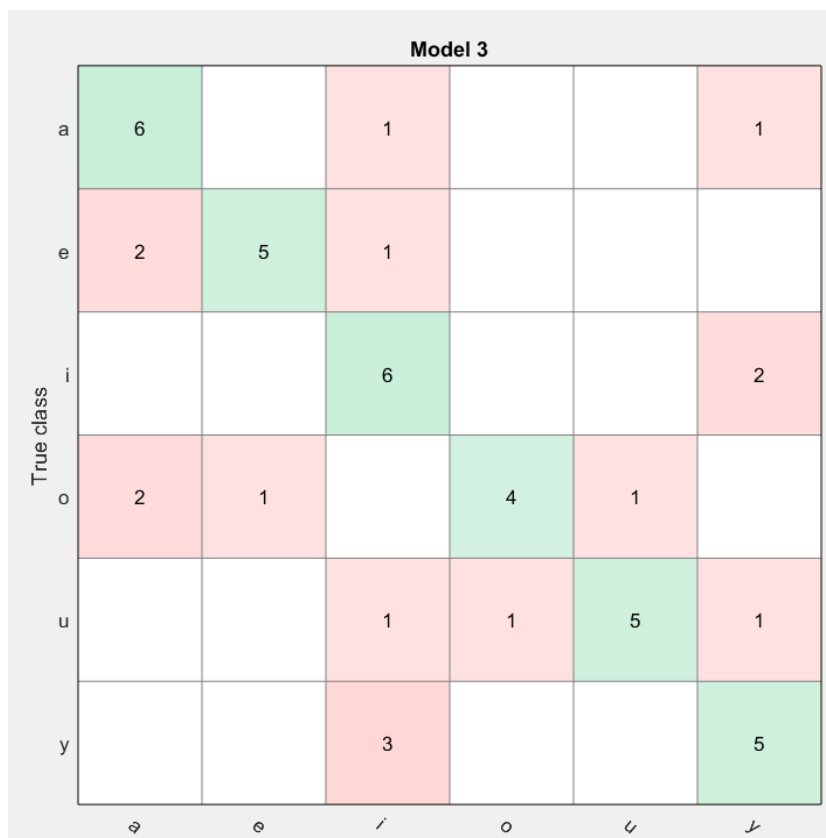


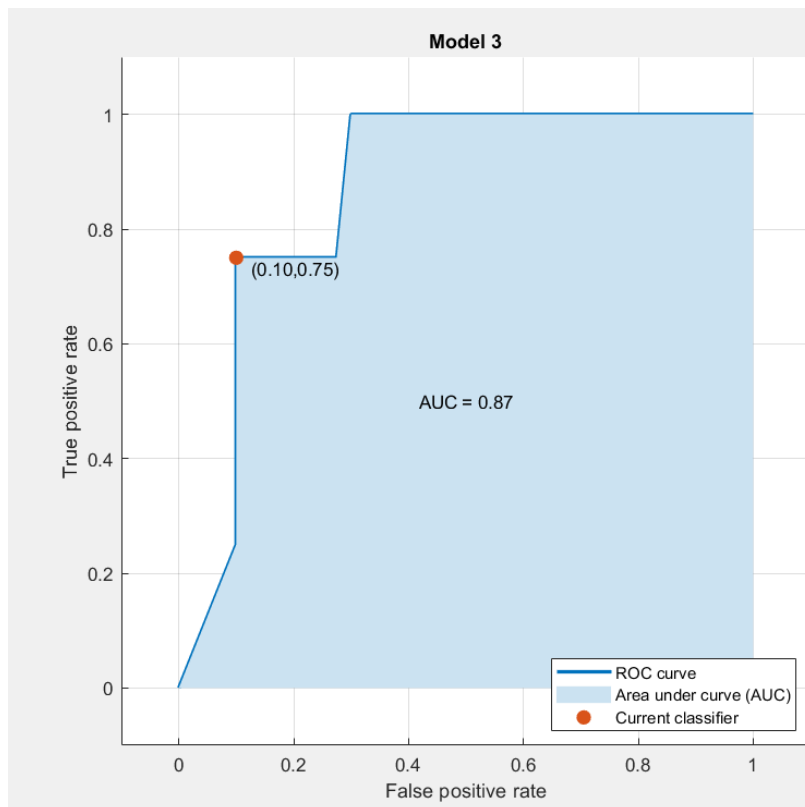
## SVM





Tree





Wyniki na zbiorze testowym dla KNN

Command Window

C =

2	1	1	0	0	0
0	2	0	1	0	1
0	0	4	0	0	0
0	0	0	3	1	0
0	0	0	0	4	0
0	1	0	0	0	3

accuracy1 =

0.7500

*fx* >>

# Analiza i wnioski

*Czy liczba wczytanych elementów klas jest podobna do siebie? Wyjaśnij dlaczego to ma znaczenie w kontekście uczenia maszynowego*

Liczba wczytanych elementów każdej z klas jest równa 12.

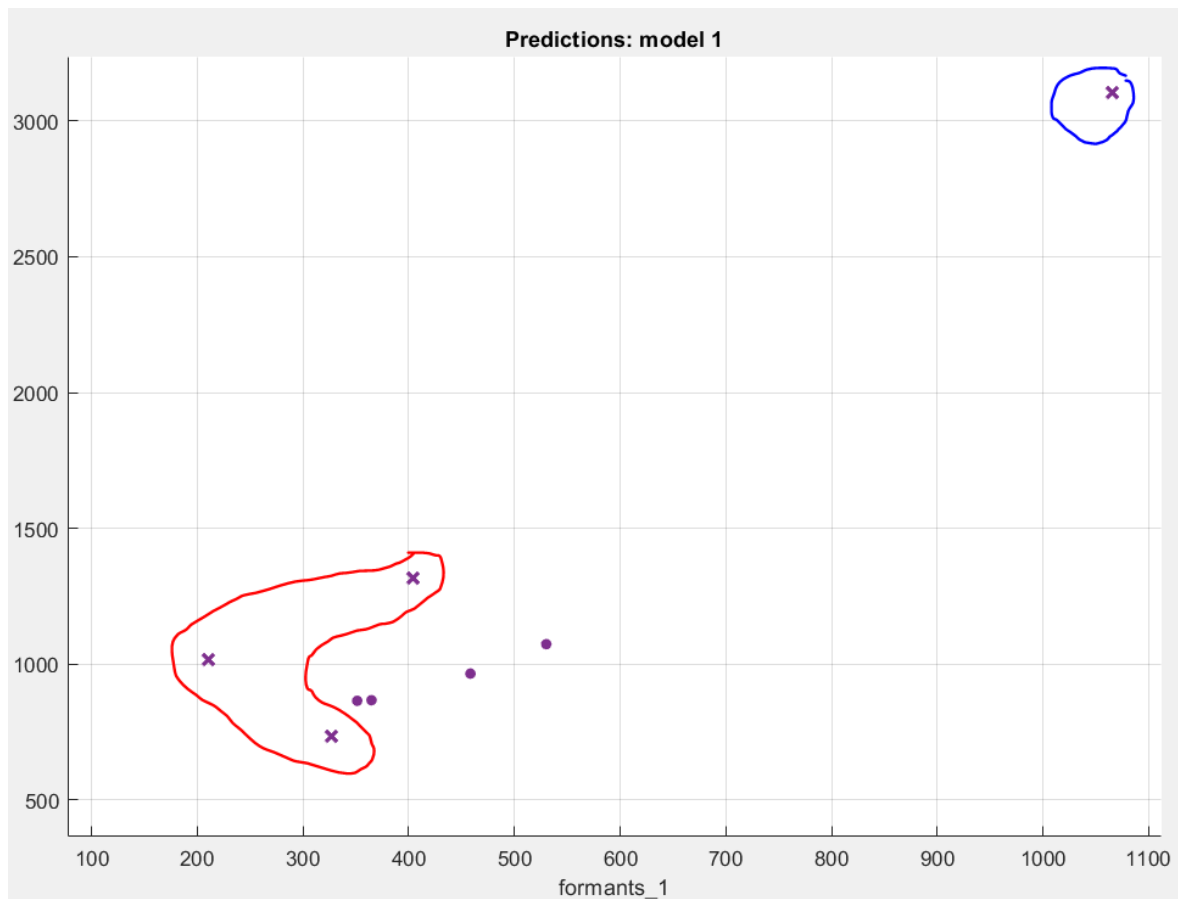
Tak, liczebność klas ma znaczenie. Gdyby jedna głoska np.: ‘a’ występowała więcej razy niż inne, to nasz model lepiej rozpoznawałby głoskę ‘a’ od innych głosek (mniej błędów w macierzy pomyłek). Jest to w pewnym sensie intuicyjne – im więcej model (mówiąc kolokwialnie) “coś przecwicz” – tym będzie w tym lepszy.

Balansowanie danych (równanie liczebności klas) nie jest zawsze potrzebne. Zależy to w głównej mierze od problemu jaki chcemy rozwiązać i do czego dany model będzie używany.

*Wyjaśnij znaczenie elementów w macierzy pomyłek. Jakie klasy są ze sobą najczęściej mylone?*

W macierzy pomyłek oś pionowa odpowiada za faktyczną klasę, a pozioma za przewidzianą/odgadniętą przez model klasę.

Na przykładzie macierzy dla KNN (dla zbioru uczącego) – w kolumnie od głoski ‘o’ mamy: 3 w wierszu od ‘a’ (3 elementy zostały uznane za ‘o’ choć tak naprawdę to były ‘a’) oraz 1 w wierszu od ‘u’. Na wykresie widać te błędy:



Czerwoną linią zaznaczono te elementy które w faktycznie były głoskami ‘a’, a niebieską linią tę która była głoską ‘u’

Klasy najczęściej ze sobą mylone:

i&y, u&o, o&a (na podstawie macierzy pomyłek dla KNN, SVM i Tree).

## Pytania

*Na czym polega walidacja krzyżowa?*

Walidacja krzyżowa, inaczej nazywana również kroswalidacją oraz sprawdzianem krzyżowym, polega na podziale próby statystycznej (zbioru wejściowego) na zbiór uczący i zbiór testowy. Prosta walidacja to najbardziej typowy rodzaj walidacji krzyżowej, w którym podział próby na wspomniane wyżej, rozłączne zbiory dokonuje się losowo. Zwykle zbiór testowy stanowi mniej niż 1/3 próby. Następnie model jest uczony na podstawie zbioru uczącego, a poprawność modelu jest sprawdzana na podstawie zbioru testowego.

Istnieje też tzw. k-krotna walidacja krzyżowa, gdzie dzieli się zbiór wejściowy na k podzbiorów, a następnie każdy z tych k zbiorów będzie służył za zbiór testowy, a pozostałe za zbiór uczący.

*Krótko wyjaśnij jak działa klasyfikator KNN.*

Jest to jedna z ważniejszych nieparametrycznych metod klasyfikacji. W tej metodzie klasyfikowany obiekt przydzielamy do tej klasy, do której należy większość z jego k najbliższych sąsiadów.

Przykładowo jeżeli mamy daną głoskę i zauważymy, że z 4 jej najbliższych sąsiadów 3 zostało zaklasyfikowanych jako 'y', a jeden jako 'a', to daną głoskę zaklasyfikujemy jako 'y'.