



AGH

**AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE**

**AGH UNIVERSITY OF SCIENCE
AND TECHNOLOGY**

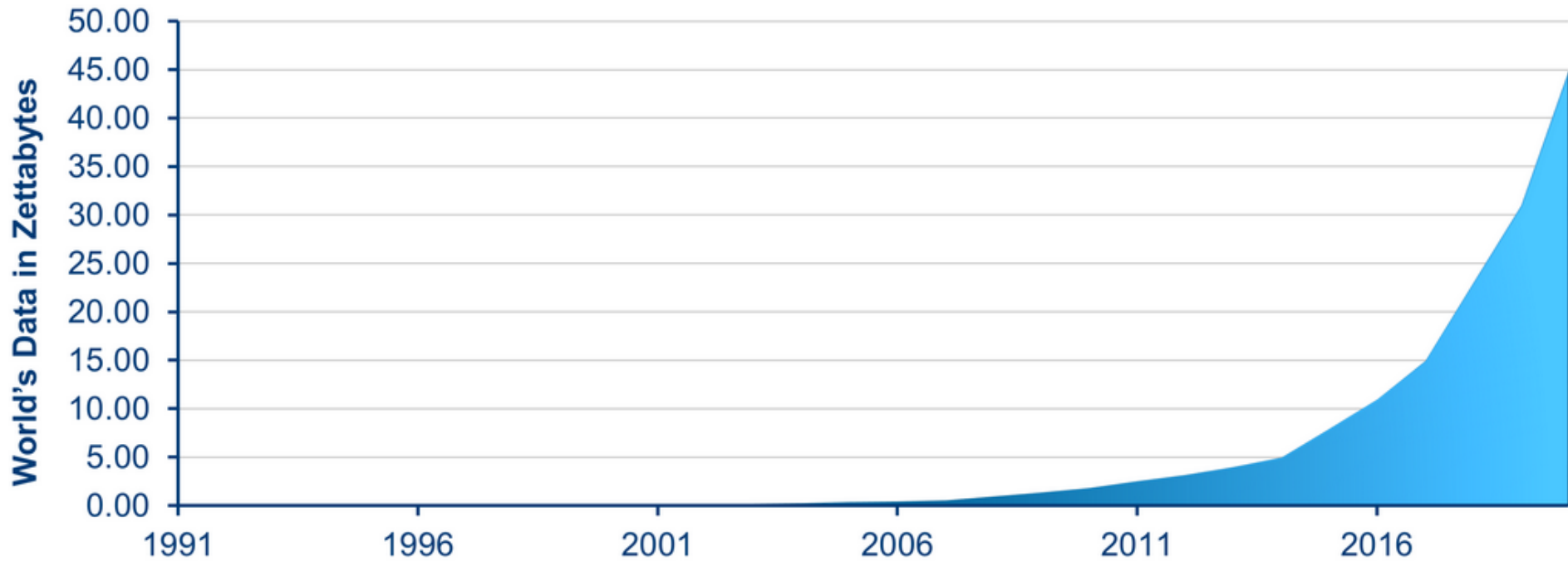
Metody Inżynierii Wiedzy

Andrzej Stach

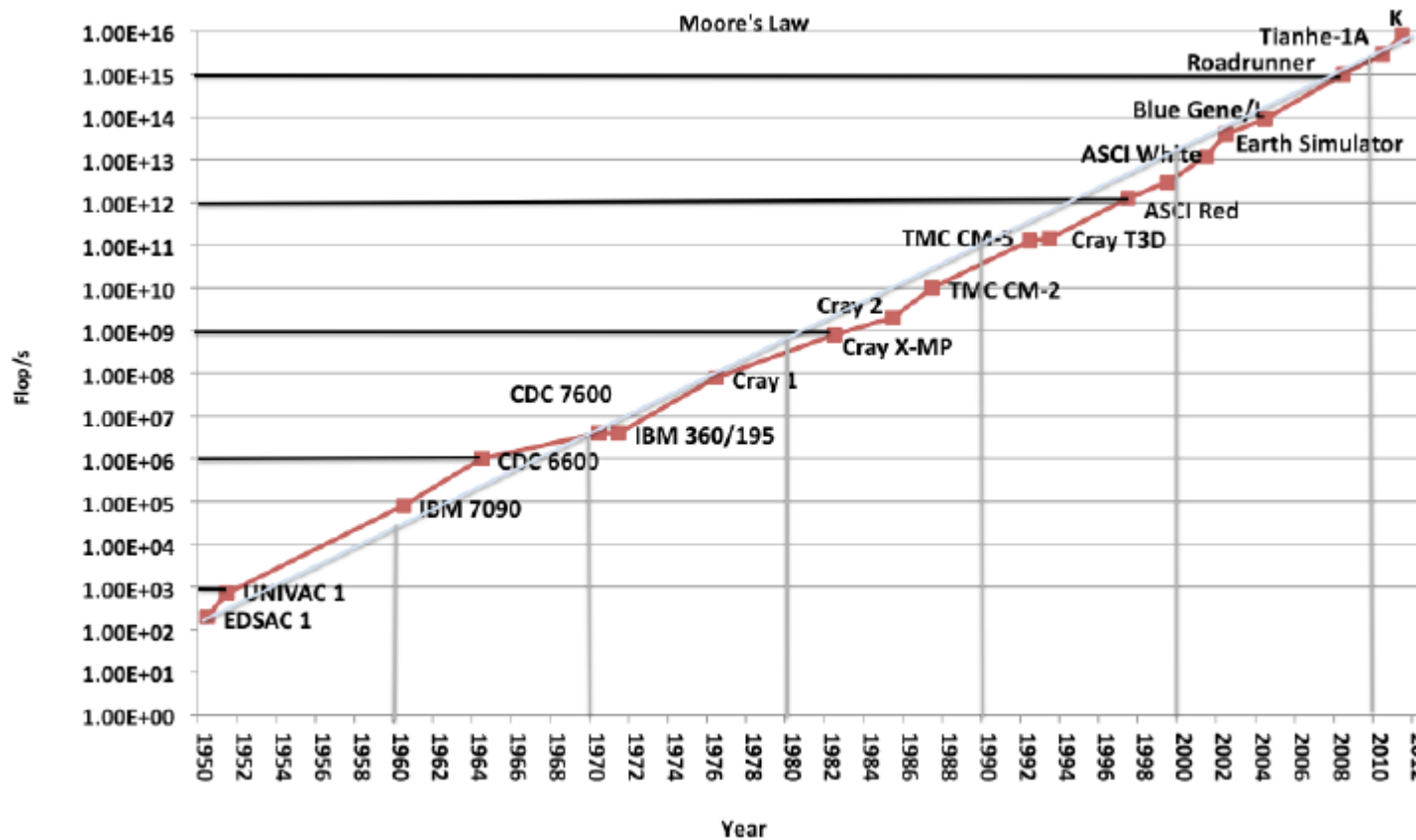
10.06.2016

World Data size

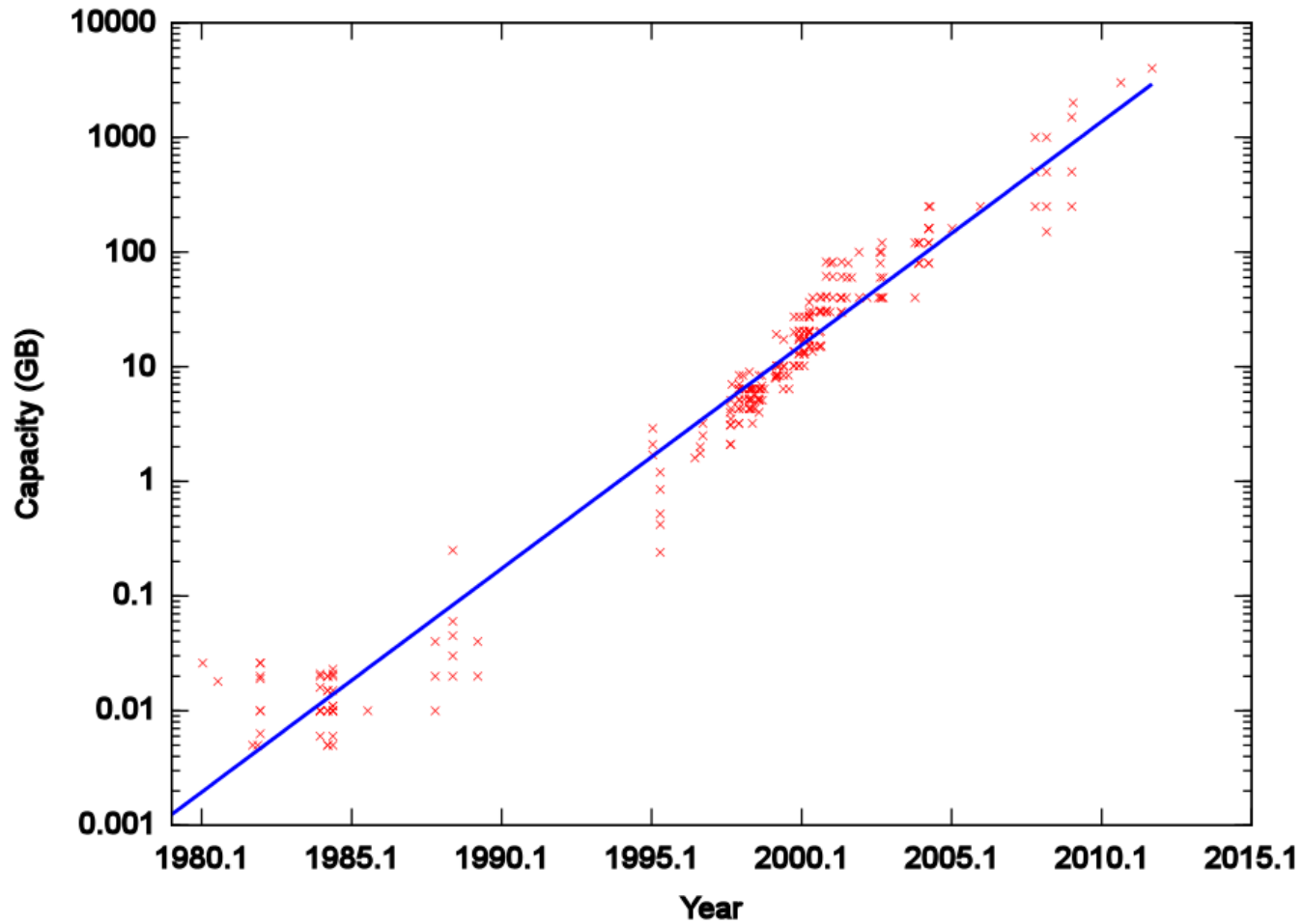
Data growth



Computer Performance



Hard drive capacity



Hard drive speed

In 2008

Typical server hard drive was **500GB** with a transfer rate of **98 MB/sec**

An entire Disk could be transferred in **85 minutes**

In 2013

Typical Server Hard Drive was **4TB** with a transfer rate of **150 MB/sec**

An entire disk could be transferred in **440 minutes**

Database

ID	Name	Account	Price	Date	Cancellation	Click to Add
1	Steven Shwimi	Gold	\$500	8/7/2009	6/30/2010	
2	Ruby jason	Premium	\$450	8/5/2009	6/1/2010	
3	Mark Dyne	Silver	\$300	8/16/2009	6/1/2010	
4	Philip namdaf	Basic	\$200	8/6/2009	6/1/2010	
5	Erik Bawn	Gold	\$500	7/15/2009	6/30/2010	
6	Ricky ben	Silver	\$300	1/13/2010	6/1/2010	
7	Miecky	Basic	\$200	8/5/2009	6/1/2010	
8	sdMcfaden	Gold	\$500	4/3/2010	6/30/2010	
9	Jack	Gold	\$500	10/15/2008	6/30/2010	
10	Philip namdaf	Silver	\$300	9/11/2008	6/1/2010	
*	(New)					

Czym jest AGDS?

AGDS to asocjacyjna struktura grafowa.

Pozwala ona na szybki dostęp do dowolnych danych, a także

reprezentowanych relacji między tymi danymi.



Przewagi AGDS nad tablicą

- » Szybszy dostęp do danych
- » Powiązanie danych przez dodatkowe relacje
- » Dane posortowane względem wszystkich atrybutów równocześnie
- » Usunięcie duplikatów

Aplikacja

AGDS

Data Table

Param	leaf-length	leaf-width	petal-length	petal-width	class	Similarit
1	5.1	3.5	1.4	0.2	Iris-setosa	0.0
2	4.9	3.0	1.4	0.2	Iris-setosa	0.0
3	4.7	3.2	1.3	0.2	Iris-setosa	0.0
4	4.6	3.1	1.5	0.2	Iris-setosa	0.0
5	5.0	3.6	1.4	0.2	Iris-setosa	0.0
6	5.4	3.9	1.7	0.4	Iris-setosa	0.0
7	4.6	3.4	1.4	0.3	Iris-setosa	0.0
8	5.0	3.4	1.5	0.2	Iris-setosa	0.0
9	4.4	2.9	1.4	0.2	Iris-setosa	0.0
10	4.9	3.1	1.5	0.1	Iris-setosa	0.0
11	5.4	3.7	1.5	0.2	Iris-setosa	0.0
12	4.8	3.4	1.6	0.2	Iris-setosa	0.0
13	4.8	3.0	1.4	0.1	Iris-setosa	0.0
14	4.3	3.0	1.1	0.1	Iris-setosa	0.0
15	5.8	4.0	1.2	0.2	Iris-setosa	0.0
16	5.7	4.4	1.5	0.4	Iris-setosa	0.0
17	5.4	3.9	1.3	0.4	Iris-setosa	0.0
18	5.1	3.5	1.4	0.3	Iris-setosa	0.0
19	5.7	3.8	1.7	0.3	Iris-setosa	0.0
20	5.1	3.8	1.5	0.3	Iris-setosa	0.0
21	5.4	3.4	1.7	0.2	Iris-setosa	0.0

Parameters:

attributes:

Number of neighbours:

Last Operation Time: / Errors:

Attribute to sort:

MODE:

Performance

» Tworzenie grafu AGDS (czas w uS)

Próba Dane	1	2	3	4	5	średnia
Iris (150 rekordów)	4 665	1 397	2 765	1 754	3 377	2 791,6
Wine (178 rekordów)	13 814	3 397	3 725	3 692	3 100	5 545,6
Iris extended (1599 rekordów)	23 889	15 959	16 164	15 348	22 266	18 725,2

Performance

» Sortowanie po atrybucie (czas w uS)

Próba Dane	1	2	3	4	5	średnia
Iris (tabela)	129	97	88	526	132	194,4
Iris (AGDS)	89	73	68	68	66	72,8
Wine (tabela)	276	350	261	200	143	246
Wine (AGDS)	112	82	97	116	94	100,2
Iris Large (tabela)	1 827	821	7 381	457	733	2 243,8
Iris Large (AGDS)	809	699	704	893	911	803,2

Performance

» Obliczanie podobieństwa (czas w uS)

Próba Dane	1	2	3	4	5	średnia
Iris (tabela)	808	461	288	329	281	433,4
Iris (AGDS)	305	291	209	158	218	236,2
Wine (tabela)	2 625	795	862	1 251	951	1296,8
Wine (AGDS)	1 135	634	479	674	409	666,2
Iris Large (tabela)	11 523	7 550	4 341	3 649	2 905	5 993,6
Iris Large (AGDS)	6 051	18 813	21 709	4 402	2 744	10 743,8

Performance

» Filtracja (czas w uS)

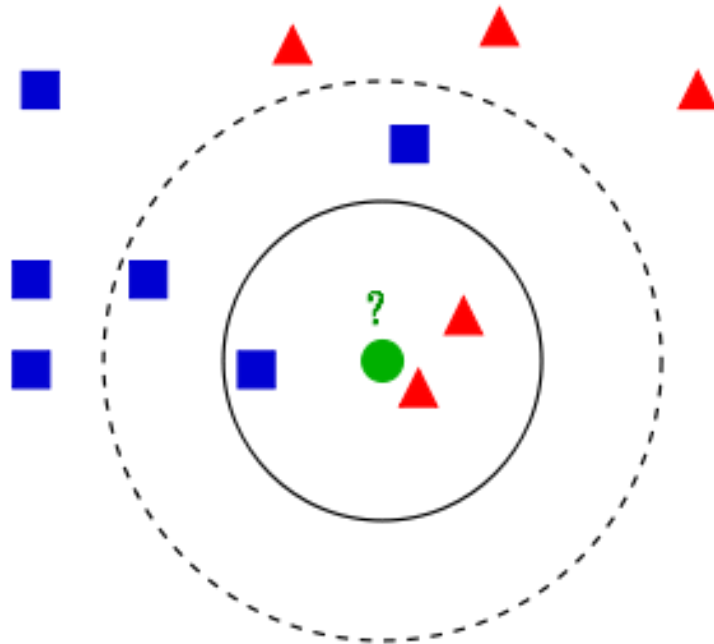
Próba Dane	1	2	3	4	5	średnia
Iris (tabela)	318	342	364	402	364	358
Iris (AGDS)	147	131	123	132	125	131,6
Wine (tabela)	1 028	1 307	997	1 039	1030	1 080,2
Wine (AGDS)	148	240	873	528	1021	562
Iris Large (tabela)	29 953	11 723	17 570	19 222	34 253	22 544,2
Iris Large (AGDS)	2 407	2 110	2 694	978	784	1 794,6



Inne aplikacje

Metoda K najbliższych sąsiadów

- Dane wczytywane z pliku
- Użytkownik ma wybór ilu sąsiadów ma być branych pod uwagę dla wytypowania klasy obiektu



Metoda K najbliższych sąsiadów

```
Wprowadź współrzędne nowego elementu:  
Leaf length:  
5.2  
Leaf width:  
3.5  
Petal length:  
1.4  
Petal width:  
0.2  
Wybierz zadanie:  
1. Metoda KNN  
2. Walidacja krzyżowa  
1  
Wprowadź ilość sąsiadów:  
10  
Ilość wystąpień:  
Iris-versicolor: 0  
Iris-virginica: 0  
Iris-setosa: 10  
Dopasowany wzorzec: Iris-setosa
```


Eksploracja danych

Reguły asocjacyjne:

$$X \rightarrow Y(s, c)$$

s – wsparcie (p, że transakcja zawiera $X \cup Y$ względem wszystkich transakcji)

c – pewność (p, że transakcja zawierająca X zawiera też Y)

- Dane wczytywane z pliku
- Możliwość ustalania progu
- Przykłady reguł asocjacyjnych:

CUKIER->JAJKA (66.0%, 50.0%)

KAWA->MLEKO (77.0%, 25.0%)

MIOD->CHLEB (66.0%, 0.0%)

ID TRANSAKЦИИ	ELEMENTY TRANSAKЦИИ
1	kawa, mleko, cukier, orzeszki
2	kawa, cukier, jajka
3	kawa, chleb, cukier, masło
4	orzeszki, cukier, miód, jajka
5	masło, mleko, jajka

Walidacja krzyżowa KNN

Zbiór danych dzielony na k równolicznych zbiorów, z czego $k-1$ jest wykorzystywanych do uczenia, a jeden do walidacji. Ważne jest pomieszczenie zbioru danych wejściowych!

- Dane wczytywane z pliku
- Możliwość wpisania maksymalnej liczby podziałów

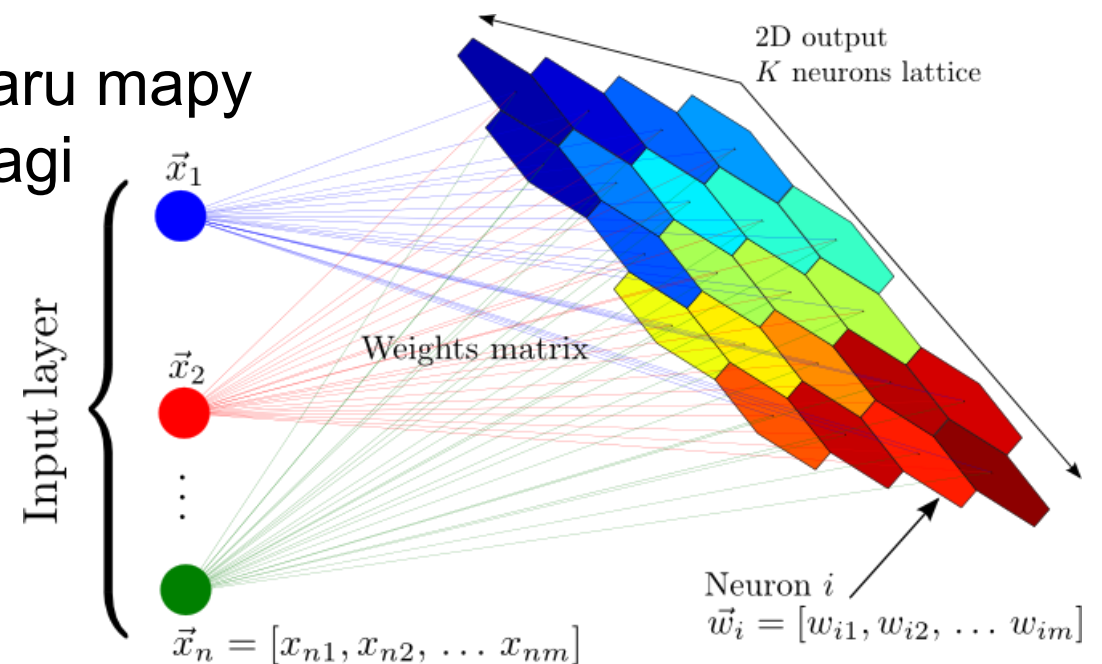
Walidacja krzyżowa KNN

```
Wprowadź współrzędne nowego elementu:  
Leaf length:  
5.2  
Leaf width:  
3.5  
Petal length:  
1.4  
Petal width:  
0.2  
Wybierz zadanie:  
1. Metoda KNN  
2. Walidacja krzyżowa  
2  
Wprowadź maksymalną ilość podziałów (mniejszą od 10)  
10  
3 0  
4 0  
4 1  
5 1  
5 2  
4 4  
5 4  
3 7  
Best k:4
```

Neuronowe mapy samoorganizujące się

SOM to sieć neuronowa ucząca się w trybie bez nauczyciela. Umożliwia ona reprezentację wielowymiarowych danych w przestrzeni o mniejszym wymiarze.

- Dane wczytywane z pliku
- Możliwość określenia rozmiaru mapy
- Możliwość wyboru funkcji wagi



Neuronowe mapy samorganizujące się

```
Podaj rozmiar mapy
```

```
5
```

```
7007
```

```
Iris-setosa Iris-versicolor Iris-virginica Iris-versicolor Iris-versicolor
```

```
Iris-virginica Iris-setosa Iris-virginica Iris-virginica Iris-versicolor
```

```
Iris-setosa Iris-versicolor Iris-versicolor Iris-versicolor Iris-virginica
```

```
Iris-versicolor Iris-versicolor Iris-versicolor Iris-versicolor Iris-versicolor
```

```
Iris-setosa Iris-versicolor Iris-versicolor Iris-versicolor Iris-versicolor
```



Dziękuję za uwagę