



WYDZIAŁ ZARZĄDZANIA
I EKONOMII

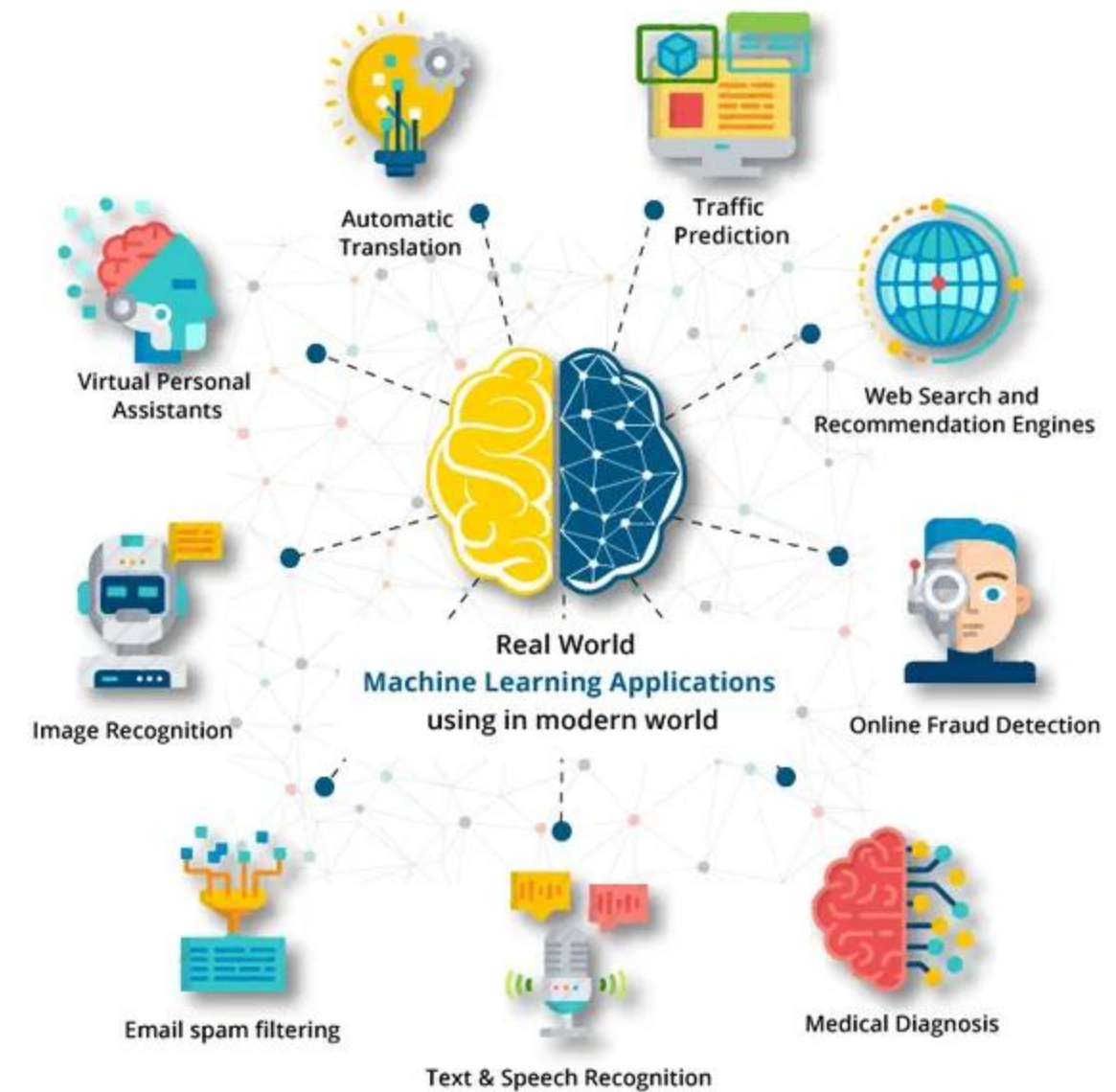


Uczenie maszynowe w R

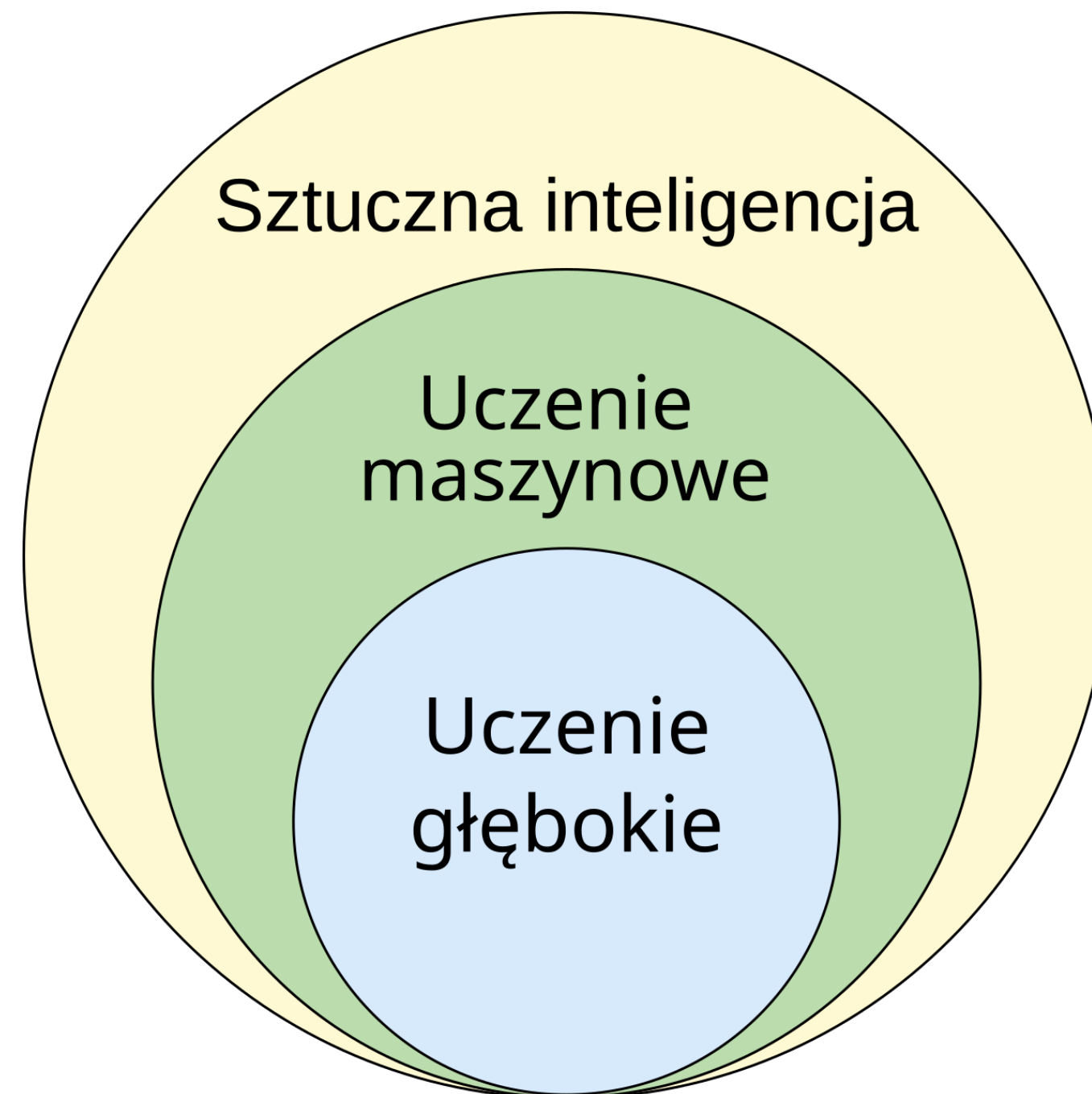
Zaglądamy do czarnej skrzynki

Uczenie maszynowe

- Uczenie maszynowe to dziedzina łącząca informatykę i matematykę, w której maszyny za pomocą odpowiednio zaprojektowanych przez ludzi algorytmów są w stanie samodzielnie "uczyć się" na podstawie dostarczonych im zbiorów danych.
- Stopień zawansowania niektórych algorytmów sprawia, że trudno zrozumieć jak dokładnie działają, stąd określenie "czarna skrzynka".



Uczenie maszynowe a AI



Język Programowania

- Język programowania, stworzony do obliczeń statystycznych, analizy zbiorów danych oraz zaawansowanej wizualizacji.
- Powszechnie stosowany wśród naukowców i statystyków



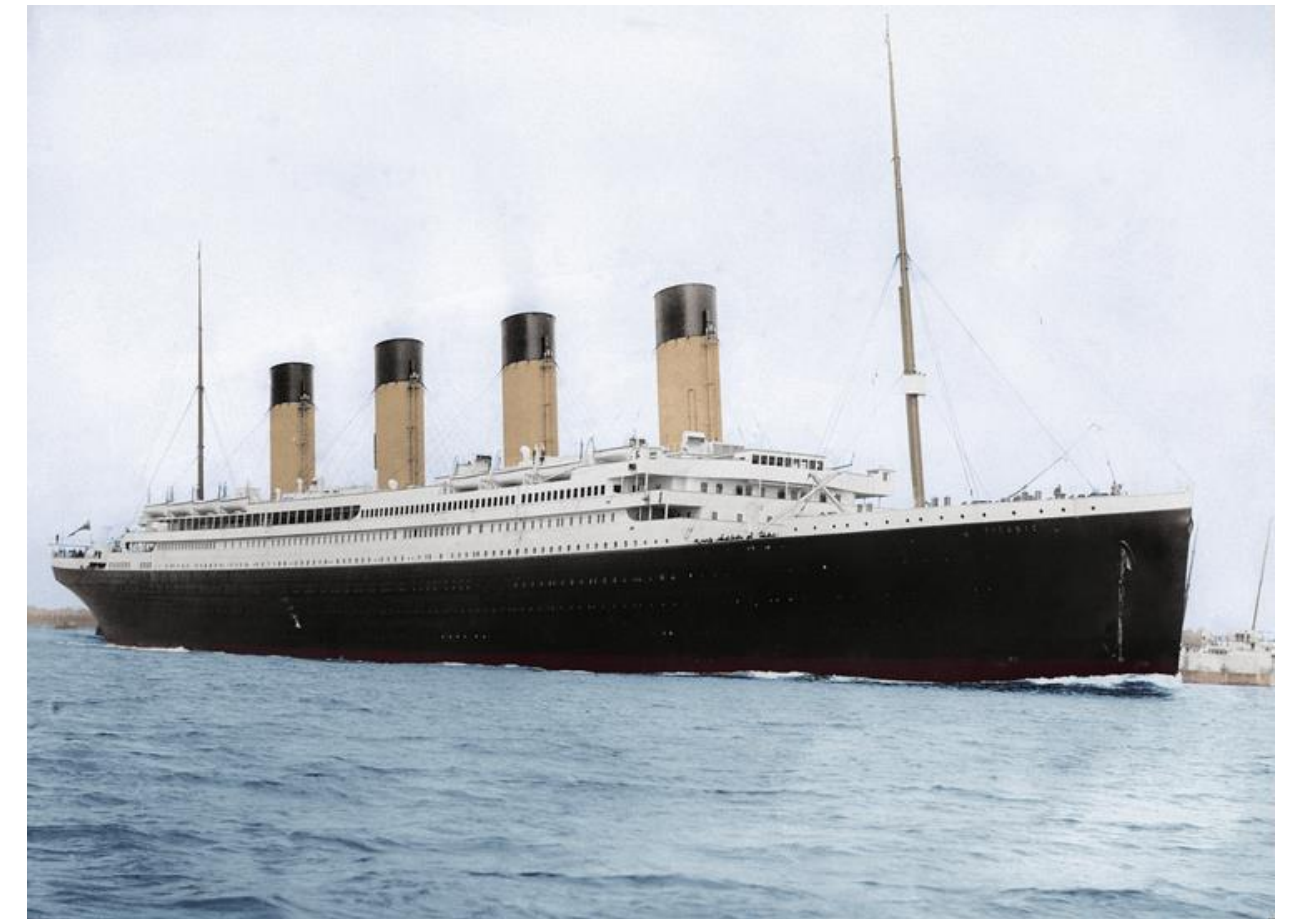
Środowisko

- Program do tworzenia, edytowania i korzystania z plików napisanych w języku R.



RMS Titanic

- Był to największy i najbardziej luksusowy statek pasażerski swoich czasów, wyruszył w dziewiczy rejs z Southampton do Nowego Jorku 10 kwietnia 1912 roku. Niestety, 14 kwietnia statek zderzył się z górą lodową na północnym Atlantyku, co spowodowało poważne uszkodzenia i zatonięcie statku kilka godzin później. Tragedia ta pochłonęła życie ponad 1500 osób i stała się jednym z najbardziej tragicznych wydarzeń w historii morskiej.



Zestaw danych

Zestaw danych, który analizujemy, zawiera informacje o ponad 2200 pasażerach i członkach załogi Titanica.

Dane obejmują:

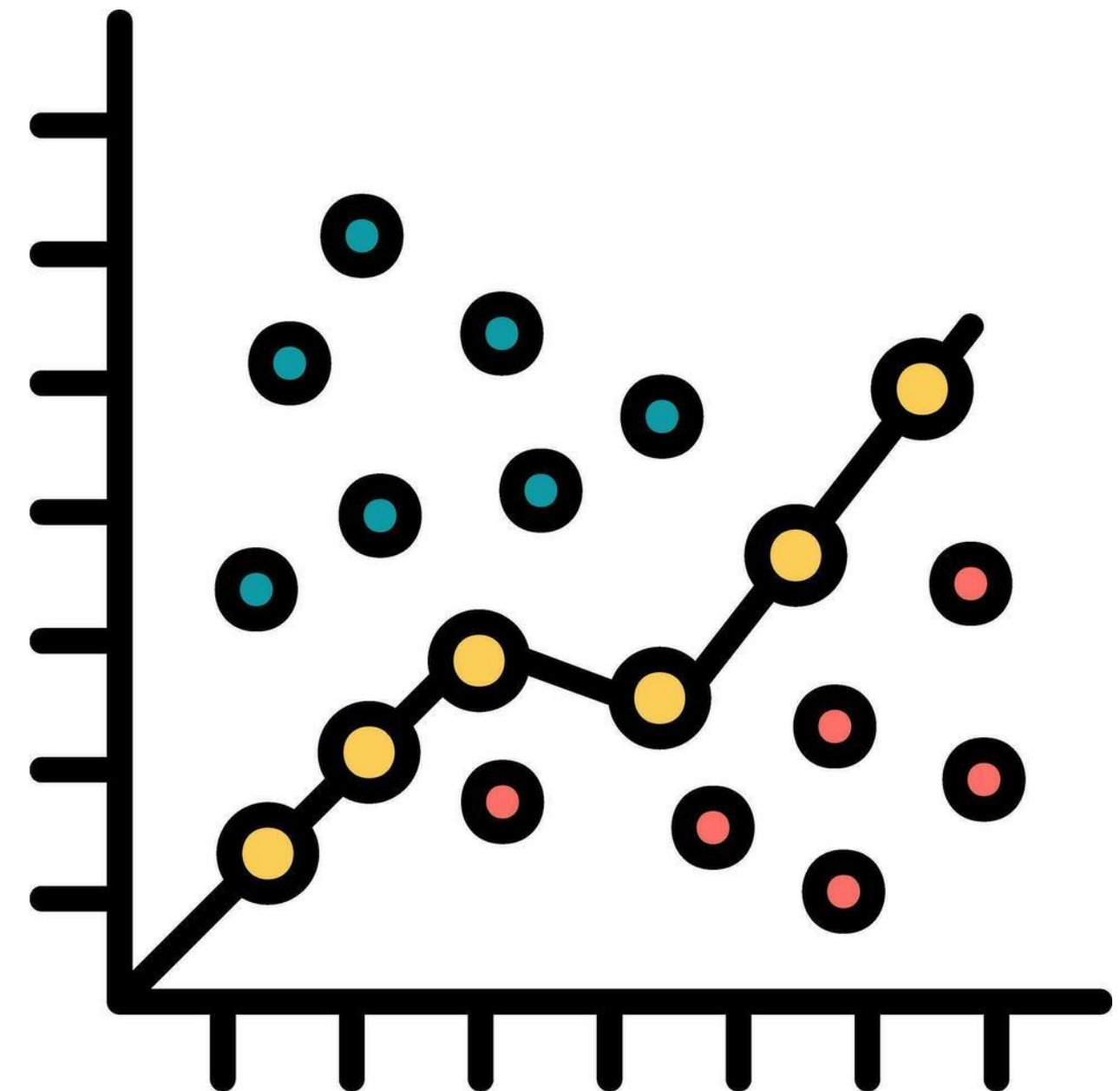
- wiek, płeć, klasę podróży
- miejsce, z którego zaokrętowano
- kraj pochodzenia
- Liczbę bliskich osób na pokładzie:
 - *sibsp* – liczba rodzeństwa i/lub małżonków na pokładzie
 - *parch* – liczba rodziców i/lub dzieci na pokładzie
- opłatę za bilet
- czy dana osoba przeżyła katastrofę (Najważniejsze!)

	gender	age	class	embarked	country	fare	sibsp	parch	survived
1	male	42.0000000	3rd	Southampton	United States	7.1100	0	0	no
2	male	13.0000000	3rd	Southampton	United States	20.0500	0	2	no
3	male	16.0000000	3rd	Southampton	United States	20.0500	1	1	no
4	female	39.0000000	3rd	Southampton	England	20.0500	1	1	yes
5	female	16.0000000	3rd	Southampton	Norway	7.1300	0	0	yes
6	male	25.0000000	3rd	Southampton	United States	7.1300	0	0	yes
7	male	30.0000000	2nd	Cherbourg	France	24.0000	1	0	no
8	female	28.0000000	2nd	Cherbourg	France	24.0000	1	0	yes
9	male	27.0000000	3rd	Cherbourg	Lebanon	18.1509	0	0	yes
10	male	20.0000000	3rd	Southampton	Finland	7.1806	0	0	yes
11	male	30.0000000	3rd	Southampton	Sweden	7.0500	0	0	no
12	male	27.0000000	3rd	Southampton	England	8.0100	0	0	no
13	female	40.0000000	3rd	Southampton	Sweden	9.0906	1	0	no
14	male	0.8333333	3rd	Southampton	England	9.0700	0	1	yes
15	female	18.0000000	3rd	Southampton	England	9.0700	0	1	yes
16	male	35.0000000	2nd	Southampton	England	13.0000	0	0	no
17	male	25.0000000	3rd	Southampton	England	7.1709	0	0	no
18	male	19.0000000	3rd	Southampton	Finland	7.1806	0	0	no

Celem analizy jest odkrycie, jakie czynniki mogły wpłynąć na przeżycie pasażerów i zrozumienie, jak wyglądała struktura społeczna na pokładzie Titanica.

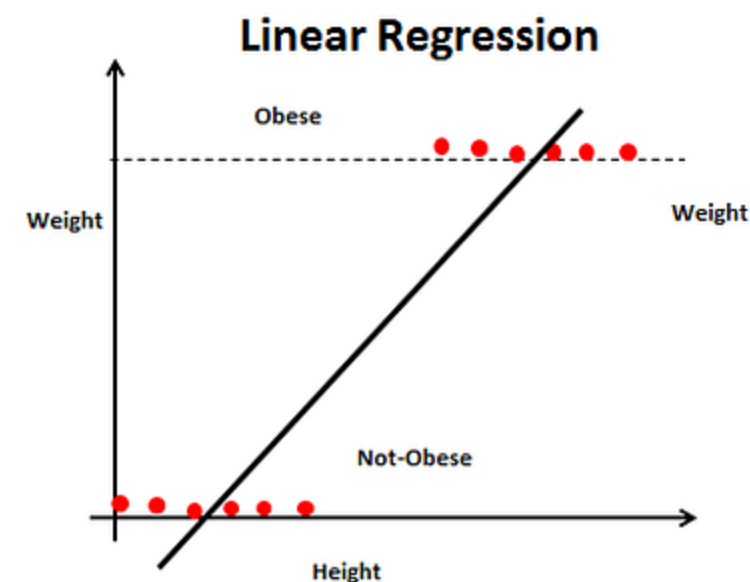
Modele, z których skorzystamy

- Model regresji liniowej
- Model regresji logistycznej
- Model lasów losowych
- Model gradient boosting
- Model maszyny wektorów nośnych (SVM)



Model regresji liniowej

Przewiduje wartość liczbową zmiennej zależnej na podstawie jednej lub więcej zmiennych niezależnych. Zakłada istnienie liniowej zależności między tymi zmiennymi. Na przykład kiedy na podstawie wieku i klasy pasażera chcemy oszacować cenę biletu.



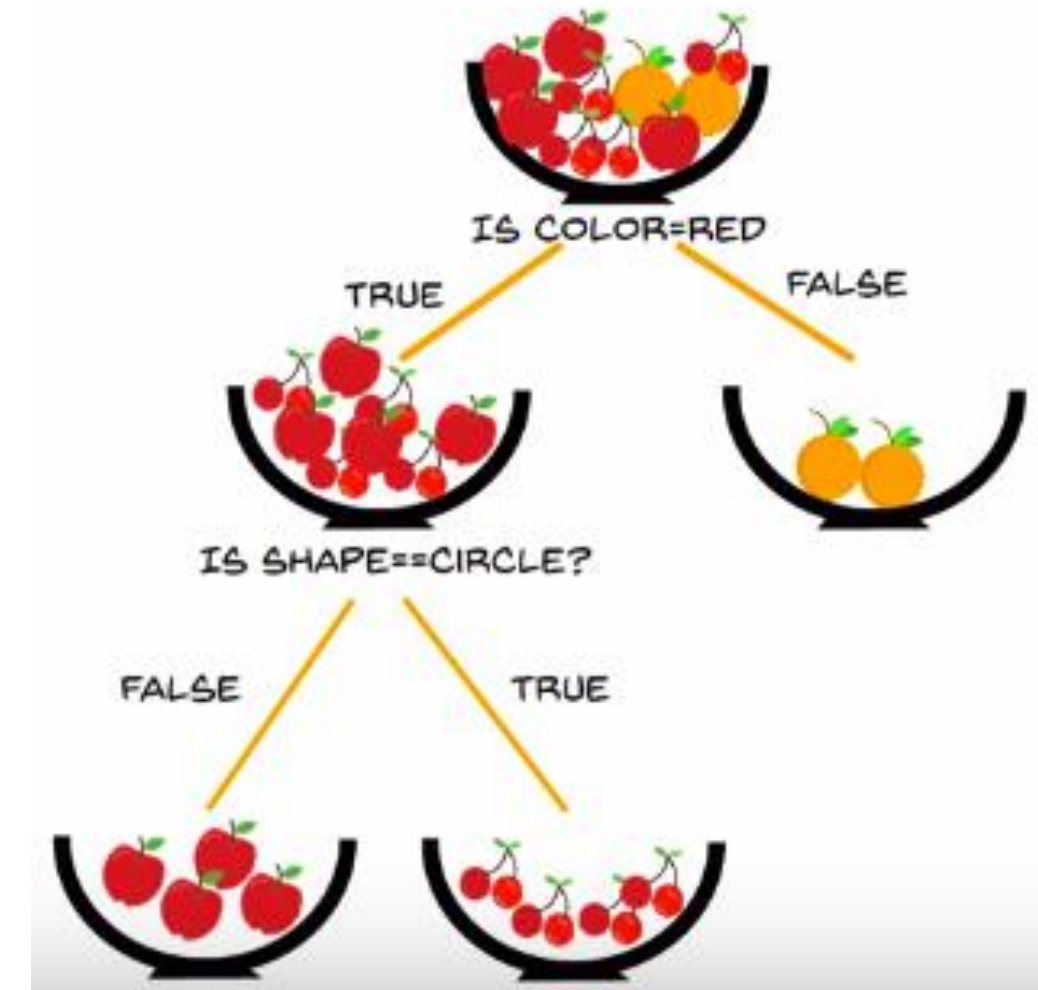
Model regresji logistycznej

Służy do przewidywania prawdopodobieństwa wystąpienia konkretnego zdarzenia, które ma tylko dwie możliwości (np. tak/nie, przeżył/ nie przeżył). Zamiast prostej linii (jak w regresji liniowej), używa krzywej w kształcie litery S, która dobrze oddziela dwie grupy.



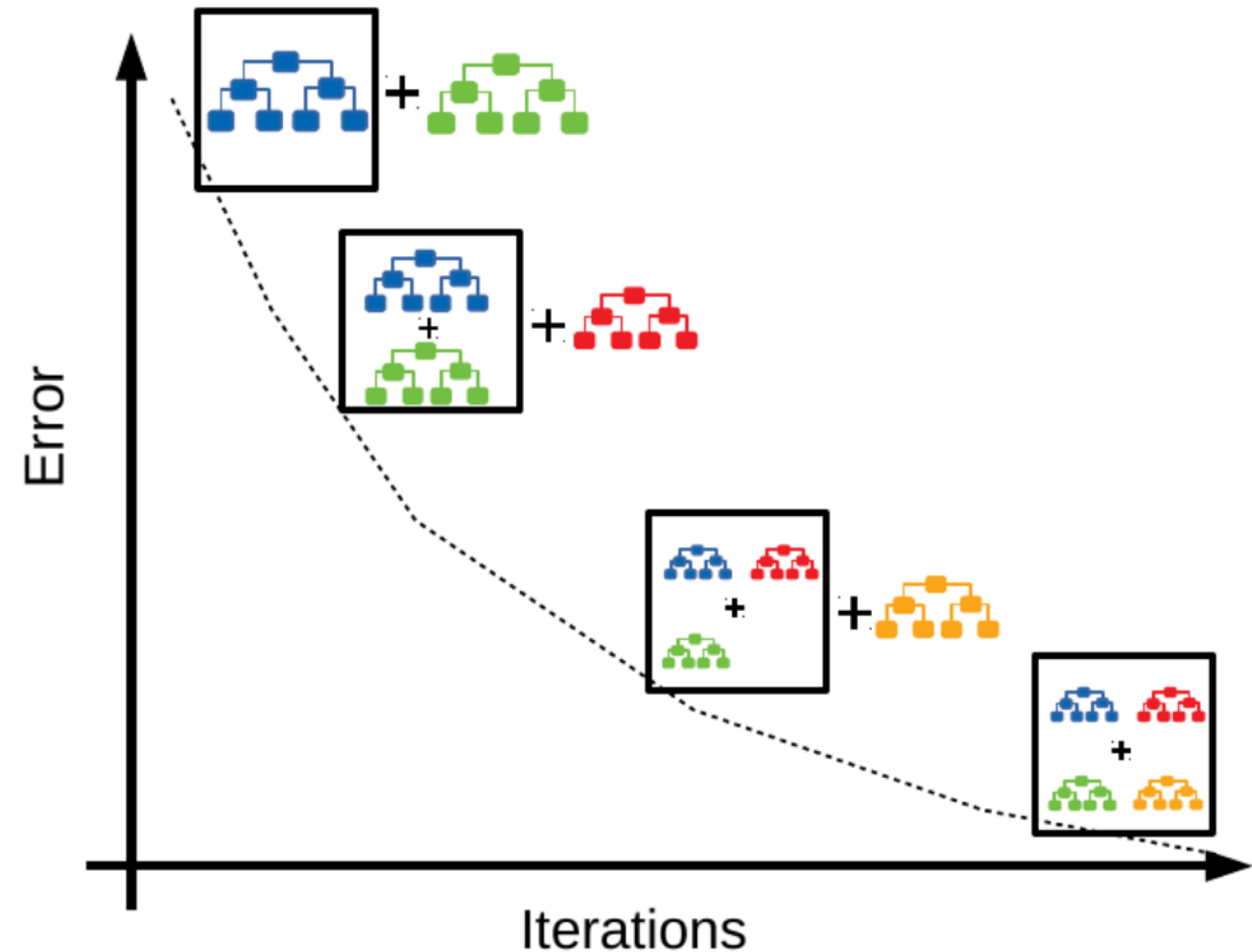
Model lasów losowych – Random Forest

Model lasów losowych to wszechstronny algorytm oparty na metodzie zespołowej. Buduje on wiele niezależnych drzew decyzyjnych na podstawie losowych podzbiorów danych, a następnie łączy ich wyniki, co zapewnia wysoką precyzję.



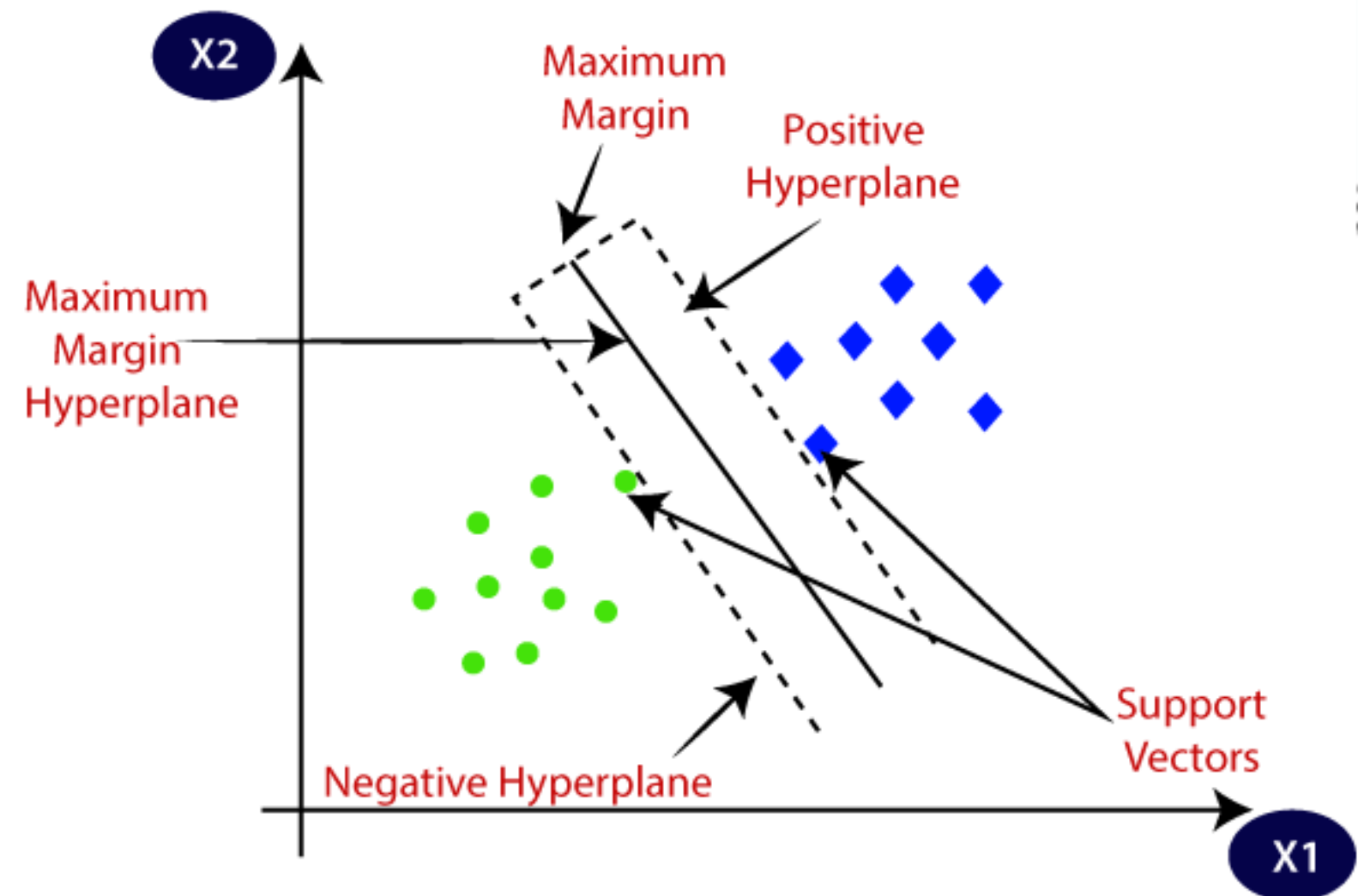
Model gradient boosting

Polega na stopniowym dodawaniu nowych, prostych modeli (często drzew decyzyjnych) do zestawu już istniejących, aby coraz lepiej przewidywać wyniki, poprawiając błędy krok po kroku.



Model maszyny wektorów (SVM)

Szuka najlepszej granicy oddzielającej różne klasy danych, maksymalizując odstęp (margines) między nimi – bardzo skuteczna w klasyfikacji, zwłaszcza gdy dane są dobrze rozdzielne.



Przykładowi pasażerowie

Variable	Jack	Rose
Class	3rd	1st
Gender	male	female
Age	20	17
Sibsp	0	0
Parch	0	1
Fare	approx. 10\$	approx. 850\$
Embarked	Southampton	Southampton



Teraz pora na Ciebie!

```
stworzenie_predykcji(  
  gender = "male", # wpisać jedno z: "female", "male"  
  age = 67,  
  class = "2nd", # wpisać jedno z: "1st", "2nd", "3rd", "deck crew", "engineering crew",  
    "restaurant staff", "victualling crew"  
  embarked = "Belfast", # wpisać jedno z: "Belfast", "Cherbourg", "Queenstown", "southampton"  
  fare = 100,  
  sibsp = 0,  
  parch = 2  
)
```





WYDZIAŁ ZARZĄDZANIA
I EKONOMII