

# Üzleti Elemzések Módszertana

## 2. Előadás: Osztályozás

Kuknyó Dániel  
Budapesti Gazdasági Egyetem

2023/24  
2.félév

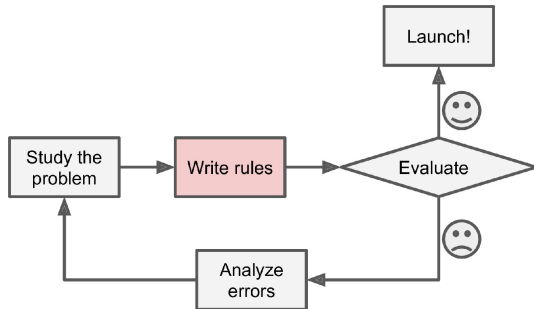
- 1 Bevezetés
- 2 Osztályozás
- 3 Osztályozás vagy regresszió?

- 1 Bevezetés
- 2 Osztályozás
- 3 Osztályozás vagy regresszió?

# A determinisztikus szemléletmód

A hagyományos szoftverfejlesztési folyamatmodell eljárása:

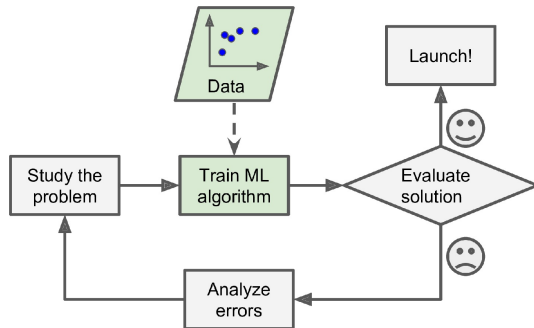
- 1 Az adott jelenség megfigyelése és adatok rögzítése
- 2 A megfigyelésekre olyan szabályok kidolgozása, amelyek jól leírják azt
- 3 A létrejött szabályrendszer kiértékelése
- 4 Rendszer fejlesztése a hibák alapján
- 5 Iteráció



# A gépi tanulás szemléletmód

A gépi tanulás szemléletének  
folyamatmodellje:

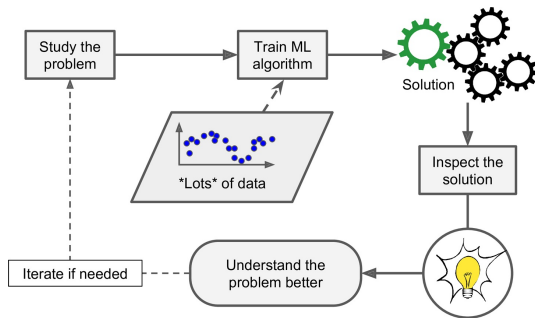
- 1 Adott jelenség megfigyelése és adatok rögzítése
- 2 Gépi tanulási modell tanítása az adatokon a szakterületi tudás segítségével
- 3 Modell kiértékelése
- 4 Hibák elemzése és kiértékelése
- 5 Iteráció



# Tanítás automatizálása adatalapúan

Az gépi tanuló modellek tanítása és kiértékelése hosszú távon egy iteratív folyamat már létező keretrendszerekkel, mint az MLOps. Ennek számos területen vannak előnyei:

- Adaptáció az új adatokhoz
- Javuló modell teljesítmény
- Hibák és problémák azonosítása
- Új technológiai fejlődés integrálása
- Skálázhatóság és rugalmasság
- Szakterületi következtetések az elemzések által

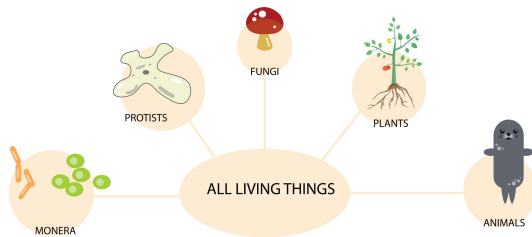


- 1 Bevezetés
- 2 **Osztályozás**
- 3 Osztályozás vagy regresszió?

# Osztályozás

## Osztályozás

Az osztályozás a felügyelt gépi tanulás egyik alapvető feladata, amelynek célja, hogy megtanuljon egy modellt vagy szabályrendszert egy adott bemeneti adat alapján **annak besorolására előre meghatározott kategóriákba vagy csoportokba.**



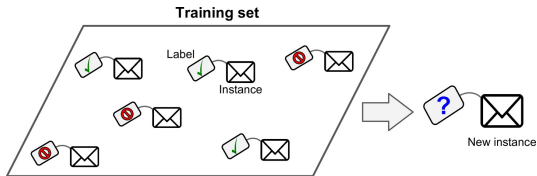
**Five Kingdom system classification**



# Modellalapú osztályozás

Az osztályozó modell feladata, hogy a tanító adathalmaza alapján olyan szabályrendszert hozzon létre, ami **képes elszeparálni egymástól az egyedeket**.

Amennyiben érkezik egy új adatpont, a modell a saját szabályrendszere segítségével már **képes lesz becslést adni annak osztályára vonatkozóan**.

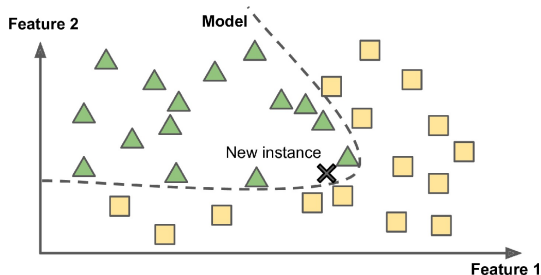


# Modellalapú osztályozás

## Döntési határ

Olyan **határérték**, amelyet a **modell állít be** az adatpontok különböző osztályokba való besorolásához.

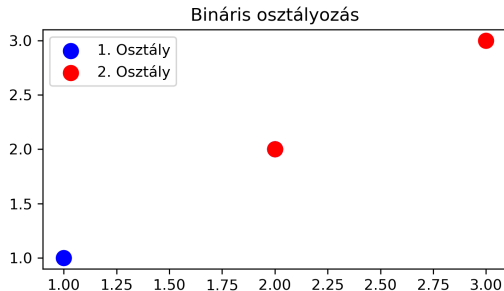
A határ **lehet egy vonal, egy sík vagy akár egy sokdimenziós felület**, attól függően, hogy milyen típusú osztályozó modellt használunk és milyen a bemeneti adatok dimenzionalitása.



# Az osztályozás fajtái

## Bináris osztályozás

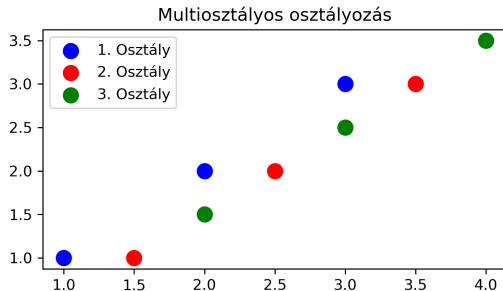
A modell két lehetséges osztály közül valamelyikbe sorolja be az egyedeket. Minden egyedhez csakis 1 osztály tartozhat.



# Az osztályozás fajtái

## Multiosztályos osztályozás

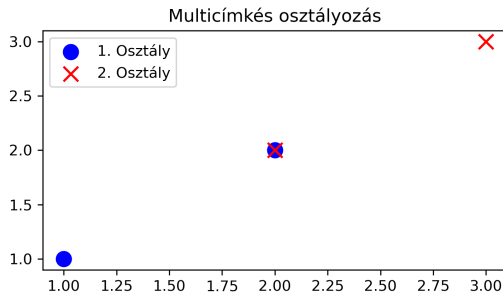
Több, mint két lehetséges kategória létezik, amibe az egyedek besorolhatók, ezek közül az egyikbe fog sorolódni az egyed. Minden egyedhez legalább és legfeljebb 1 osztály tartozik.



# Az osztályozás fajtái

## Multicímkes osztályozás

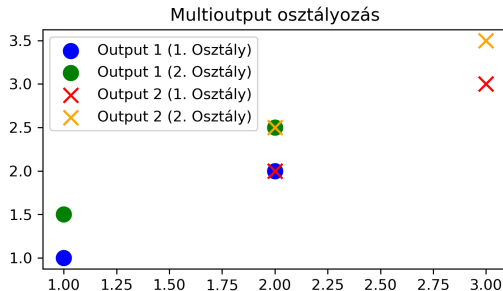
Minden mintaegyedhez több bináris vagy multicímkes címke kategóriából tartozhat osztály.



# Az osztályozás fajtái

## Multioutput osztályozás

A multicímkes osztályozás generalizált változata. Egy egyedhez egy multicímkes halmazból több elem is tartozhat.



- 1 Bevezetés
- 2 Osztályozás
- 3 Osztályozás vagy regresszió?

# Példa

A következő kis adathalmaz három sakkjátszmának rögzítette az eredményét. Minden meccs esetén rögzítésre kerültek a következő rekordok:

Rangok különbsége	Fehér nyert
200	0
-200	1
300	0