



# MILYEN FELADATOK MEGOLDÁSÁRA ALKALMAS A GÉPI TANULÁS?

Data science képzés

“A data science jó” - Bevezető

2024.09.09

Gáspár Csaba,  
CEO, data scientist



# MACHINE LEARNING

---

# ADATBÁNYÁSZATI ALAPPROBLÉMÁK

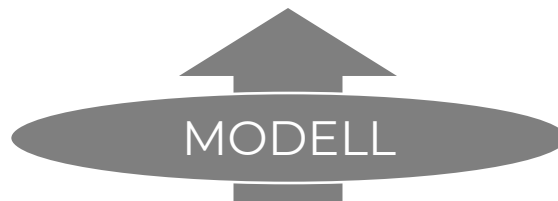
Tulajdonságleíró változók – input attribútumok

Célváltozó  
Target  
Címke

Életkor	Jövedelem	Lakásméret	Gyererek	Autók száma	Biztosítás
42	280 000	82	3	1	SIKERÜLT
88	90 000	44	0	0	NEM
22	180 000	32	0	1	SIKERÜLT
38	400 000	102	1	2	NEM

# ADATBÁNYÁSZATI ALAPPROBLÉMÁK

Életkor	Jövedelem	Lakásméret	Gyerekek	Autók száma	Biztosítás
55	180 000	62	1	3	???



Tulajdonságleíró változók – input attribútumok

Célváltozó  
Target  
Címke

Életkor	Jövedelem	Lakásméret	Gyerekek	Autók száma	Biztosítás
42	280 000	82	3	1	SIKERÜLT
88	90 000	44	0	0	NEM
22	180 000	32	0	1	SIKERÜLT
38	400 000	102	1	2	NEM

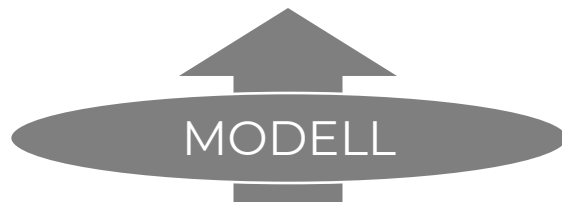
Feladat:

- Van egy tanító adathalmazom
- ... ahol ismerem a tulajdonságokat és
- ... ahol ismerem a címkéket (osztályt)
- És van egy új sor, ahol
- ... ismerem a tulajdonságokat
- ... és a modell mondja meg a címkét

# Osztályozás

Klasszifikáció  
Címkézés

Életkor	Jövedelem	Lakásméret	Gyerekek	Autók száma	Biztosítás
55	180 000	62	1	3	???



Tulajdonságleíró változók – input attribútumok

Célváltozó  
Target  
Címke

Életkor	Jövedelem	Lakásméret	Gyerekek	Autók száma	Biztosítás
42	280 000	82	3	1	SIKERÜLT
88	90 000	44	0	0	NEM
22	180 000	32	0	1	SIKERÜLT
38	400 000	102	1	2	NEM

Feladat:

- Van egy tanító adathalmazom
- ... ahol ismerem a tulajdonságokat és
- ... ahol ismerem a címkéket (osztályt)
- És van egy új sor, ahol
- ... ismerem a tulajdonságokat
- ... és a modell mondja meg a címkét

# Osztályozás

Klasszifikáció  
Címkézés

Életkor	Jövedelem	Lakásméret	Gyerekek	Autók száma	Biztosítás
55	180 000	62	1	3	???

## Alapkérdés:

Mennyire találtam el az osztályt?  
Milyen pontos volt az osztályozás?

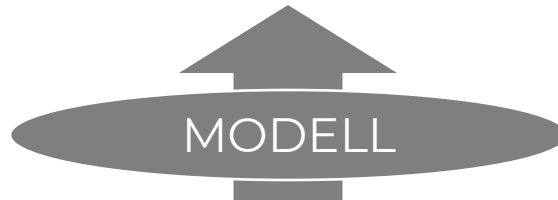
Pontosság (Accuracy) = Eltalált esetek / Összes eset

6

42	280 000	82	3	1	SIKERÜLT
88	90 000	44	0	0	NEM
22	180 000	32	0	1	SIKERÜLT
38	400 000	102	1	2	NEM

# ADATBÁNYÁSZATI ALAPPROBLÉMÁK

Életkor	Jövedelem	Lakásméret	Gyerekek	Autók száma	Lakás kora
55	180 000	62	1	3	???



Tulajdonságleíró változók – input attribútumok

Célváltozó  
Target  
Mennyiség

Életkor	Jövedelem	Lakásméret	Gyerekek	Autók száma	Lakás kora
42	280 000	82	3	1	12 év
88	90 000	44	0	0	63 év
22	180 000	32	0	1	1 év
38	400 000	102	1	2	16 év

Feladat:

- Van egy tanító adathalmazom
- ... ahol ismerem a tulajdonságokat és
- ... ahol ismerem a mennyiséget
- És van egy új sor, ahol
- ... ismerem a tulajdonságokat
- ... modell mondja meg mekkora a mennyiség

# Regresszió

Életkor	Jövedelem	Lakásméret	Gyerekek	Autók száma	Lakás kora
55	180 000	62	1	3	???

MODELL

Tulajdonságleíró változók – input attribútumok

Célváltozó  
Target  
Mennyiség

Életkor	Jövedelem	Lakásméret	Gyerekek	Autók száma	Lakás kora
42	280 000	82	3	1	12 év
88	90 000	44	0	0	63 év
22	180 000	32	0	1	1 év
38	400 000	102	1	2	16 év



Feladat:

- Van egy tanító adathalmazom
- ... ahol ismerem a tulajdonságokat és
- ... ahol ismerem a mennyiséget
- És van egy új sor, ahol
- ... ismerem a tulajdonságokat
- ... modell mondja meg mekkora a mennyiség

# Regresszió

Életkor	Jövedelem	Lakásméret	Gyerekek	Autók száma	Lakás kora
55	180 000	62	1	3	????

## Alapkérdés:

Mekkorát tévedtem?  
Átlagosán mekkora a hiba?

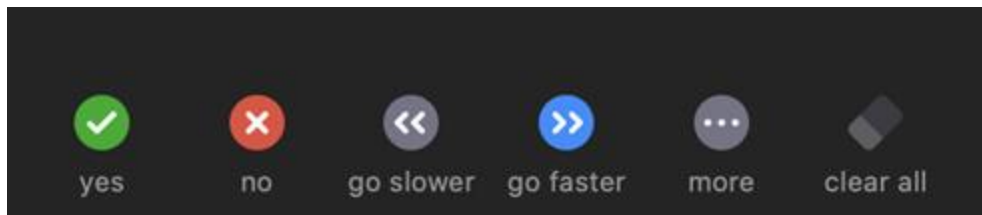
$$\text{Hiba (Error)} = |\text{Tipp} - \text{Valós érték}|$$

42	280 000	82	3	1	12 év
88	90 000	44	0	0	63 év
22	180 000	32	0	1	1 év
38	400 000	102	1	2	16 év

# TALÁLJUK KI, **MELYIK FELADAT**

---

Játszunk egy kicsit, hogy kitalálod-e,  
milyen feladatról van szó?





Churn feladat – El fog-e  
vándorolni a konkurenciához az  
ügyfél?



Kinek érdemes kiküldeni egy  
hitelkártyát?



Hány percet fog az ügyfél a  
következő hónapban  
telefonálni?



1 és 5 között milyen osztályzatot  
ad a filmre majd a felhasználó?



Többet fog-e jövőre költeni az  
ügyfél, mint 1 MFt?



Milyen körzetszámú  
mobiltelefonszáma van egy  
ügyfélnek?



# ADATBÁNYÁSZATI ALAPPROBLÉMÁK

---

## Tulajdonságleíró változók – input attribútumok

Életkor	Jövedelem	Lakásméret	Gyerekek	Autók száma
42	280 000	82	3	1
88	90 000	44	0	0
22	180 000	32	0	1
38	400 000	102	1	2

# ADATBÁNYÁSZATI ALAPPROBLÉMÁK

## Tulajdonságleíró változók – input attribútumok

Életkor	Jövedelem	Lakásméret	Gyerekek	Autók száma
42	280 000	82	3	1
88	90 000	44	0	0
22	180 000	32	0	1
38	400 000	102	1	2

MODELL

### Csoport

A. klaszter

A. klaszter

B. klaszter

B. klaszter

Feladat:

- Van egy tanító adathalmazom
- ... ahol ismerem a tulajdonságokat
- Kérdés, hogy milyen csoportokba rakható úgy, hogy
- ... csoporton belül hasonlóak
- .... csoportok ne hasonlítsanak egymásra

# Klaszterezés

Csoportképzés

## Tulajdonságleíró változók – input attribútumok

Életkor	Jövedelem	Lakásméret	Gyerekek	Autók száma
42	280 000	82	3	1
88	90 000	44	0	0
22	180 000	32	0	1
38	400 000	102	1	2

MODELL

### Csoport

A. klaszter

A. klaszter

B. klaszter

B. klaszter

# ADATBÁNYÁSZATI ALAPPROBLÉMÁK

---

## Tulajdonságleíró változók – input attribútumok

Életkor	Jövedelem	Lakásméret	Gyerekek	Autók száma
42	280 000	82	3	1
88	90 000	44	0	0
22	180 000	32	0	1
38	4 000 000	1020	1	5

# ADATBÁNYÁSZATI ALAPPROBLÉMÁK

## Tulajdonságleíró változók – input attribútumok

Életkor	Jövedelem	Lakásméret	Gyerekek	Autók száma
42	280 000	82	3	1
88	90 000	44	0	0
22	180 000	32	0	1
38	4 000 000	1020	1	5

MODELL

### Anomália

normal

normal

normal

anomália

Feladat:

- Van egy adathalmazom
- ... ahol ismerem a tulajdonságokat
- Kérdés, hogy mely elemek azok
- ... amik nagyon furcsának, nem oda illőnek tűnnek
- .... a többi elemhez képest.

# Anomália detekció

## Tulajdonságleíró változók – input attribútumok

Életkor	Jövedelem	Lakásméret	Gyerekek	Autók száma
42	280 000	82	3	1
88	90 000	44	0	0
22	180 000	32	0	1
38	4 000 000	1020	1	5

MODELL

### Anomália

normal

normal

normal

anomália

Osztályozás

Klaszterezés

Regresszió

Anomália  
detekció

Osztályozás

Klaszterezés

Regresszió

Anomália  
detekció

Domain  
specifikus eljárások



Osztályozás

Klaszterezés

Regresszió

Anomália  
detekció

Domain  
specifikus eljárások

Asszociációs  
szabályok

Python fordító - interpreter

*„... mint egy egyszerű karakteres chat-ablak”*

iPython Notebook:

*„... mint egy szép Messenger alkalmazás”*

<https://bme.dmlab.hu>

BME\_D@t@24\_osz

Alternatíva saját gépre: Anaconda (python, iPython, Spider)



# Mennyire jó a tempó Python írásakor?

*yes - pont jó*

*>> - túl gyors*

*<< - túl lassú*

# TUTORIALS

---

## **Gyorstalpaló**

<https://learnxinyminutes.com/docs/python3/>

## **Akinek új a programozás**

[https://en.wikibooks.org/wiki/Non-Programmer%27s\\_Tutorial\\_for\\_Python\\_3](https://en.wikibooks.org/wiki/Non-Programmer%27s_Tutorial_for_Python_3)

## **Learn Python on the Hard Way**

<http://learnpythonthehardway.org/book/>



ELEMENTARY

CHECKIO.ORG



HOME

60

# cyber-dojjo.org

the place to practice programming



**setup a new practice session**

**enter a practice session**

**review a practice session**

100% of your donation buys  
Raspberry Pi computers to  
help children learn to program

**please  
donate**



**Ha KÉRDÉSED VAN,  
BÁTRAN KERESS  
MINKET!**

**GÁSPÁR CSABA**  
**ALAPÍTÓ, ÜGYVEZETŐ**  
[gaspar.csaba@dmlab.hu](mailto:gaspar.csaba@dmlab.hu)  
**+36 (20) 823 4154**

