



Nemzeti Adó-
és Vámhivatal

Rejtett minták, nagy hálózatok

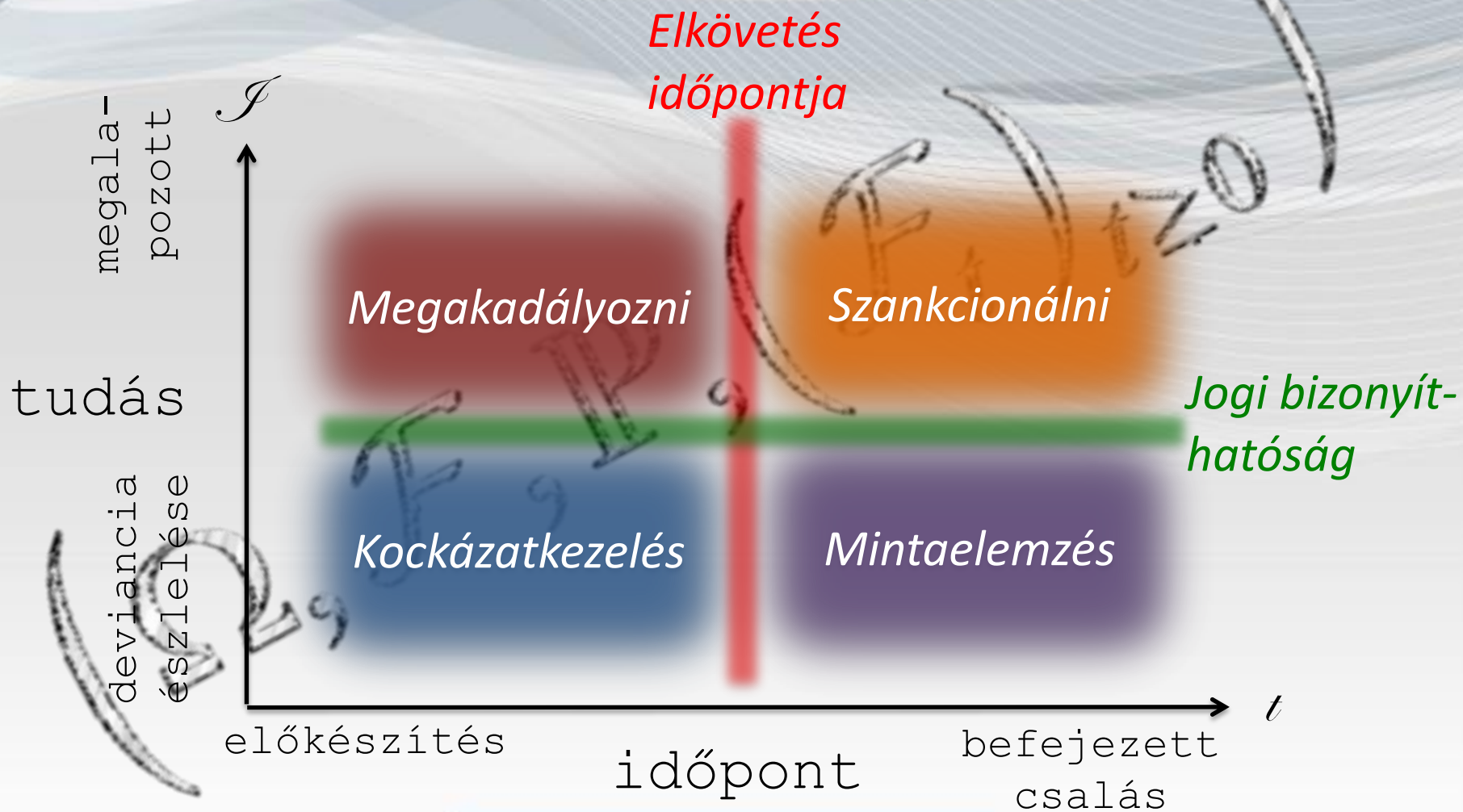
*Vágujhelyi Ferenc
informatikai elnökhelyettes*

Mi a csalás?

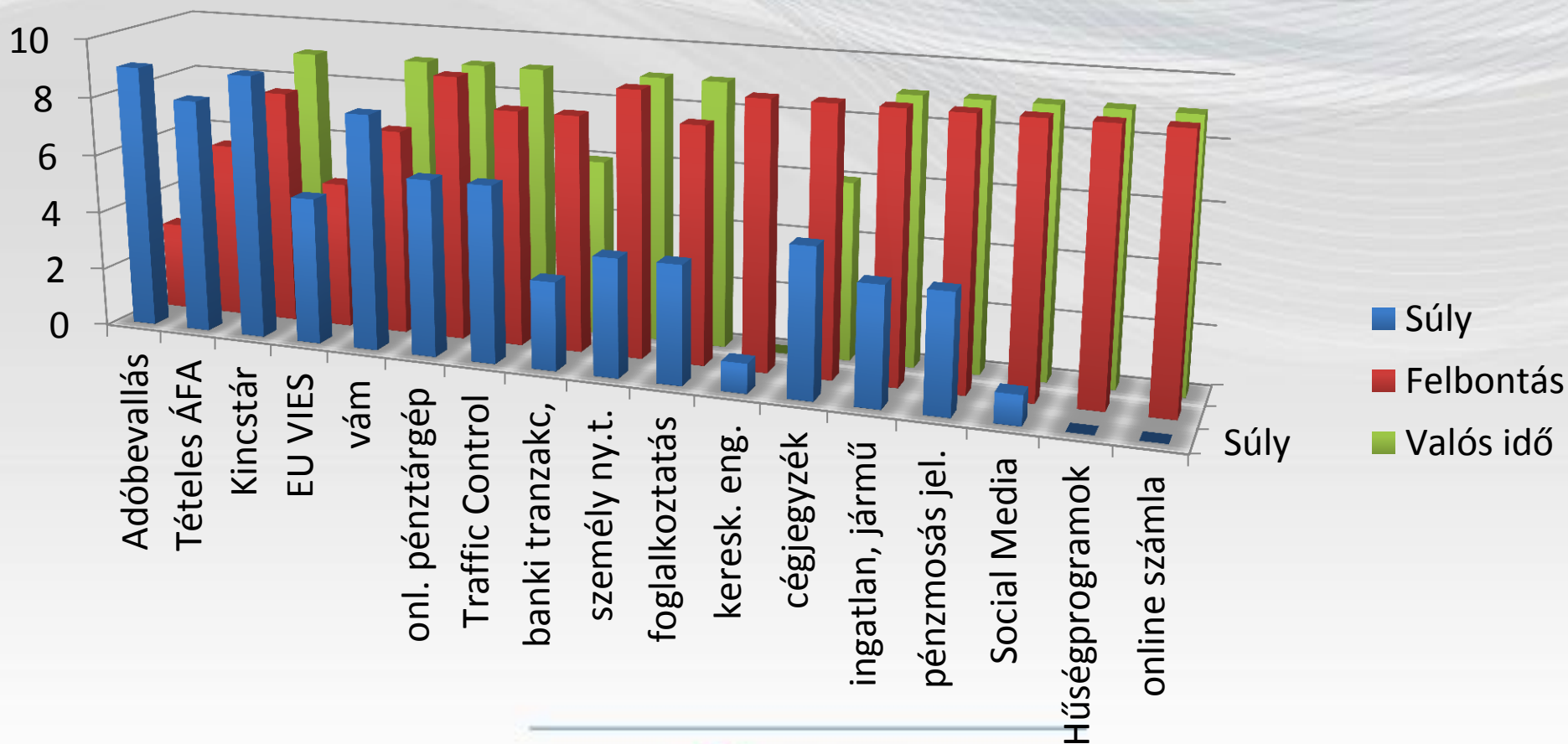
{#define NOTION_FRAUD}

- a költségvetés kárára,
- anyagi haszonért végzett,
- törvénytelen,
- nem általánosan gyakorolt (nem gyakori),
- jól átgondolt (megtervezett),
- észrevétlenül végrehajtott (rejtett),
- időben kibontakozó (jellemző az időbeli viselkedése)
- általában alaposan megszervezett bűncselekmény.

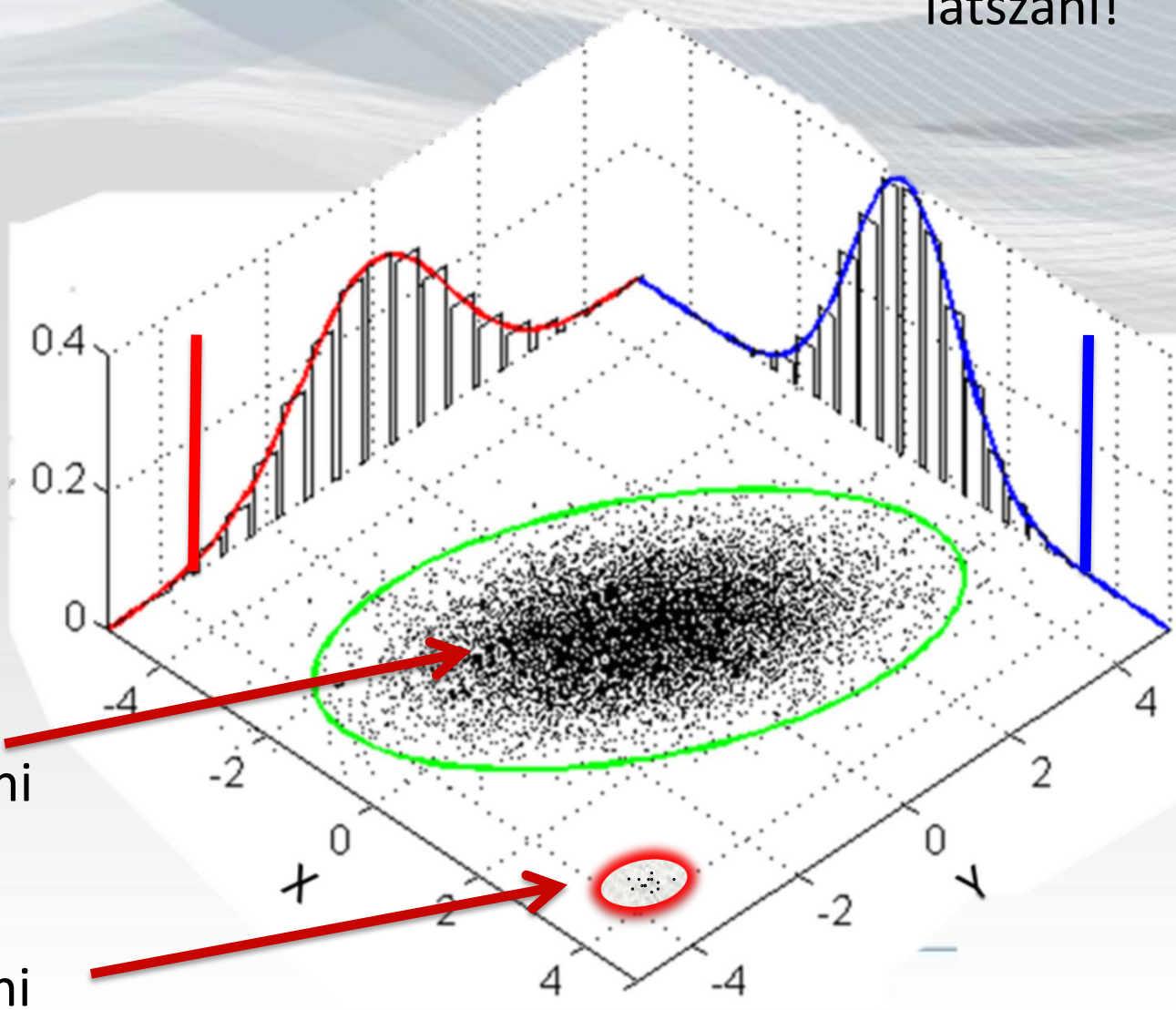
Mi a felderítés?



Miből dolgozunk? (példák)



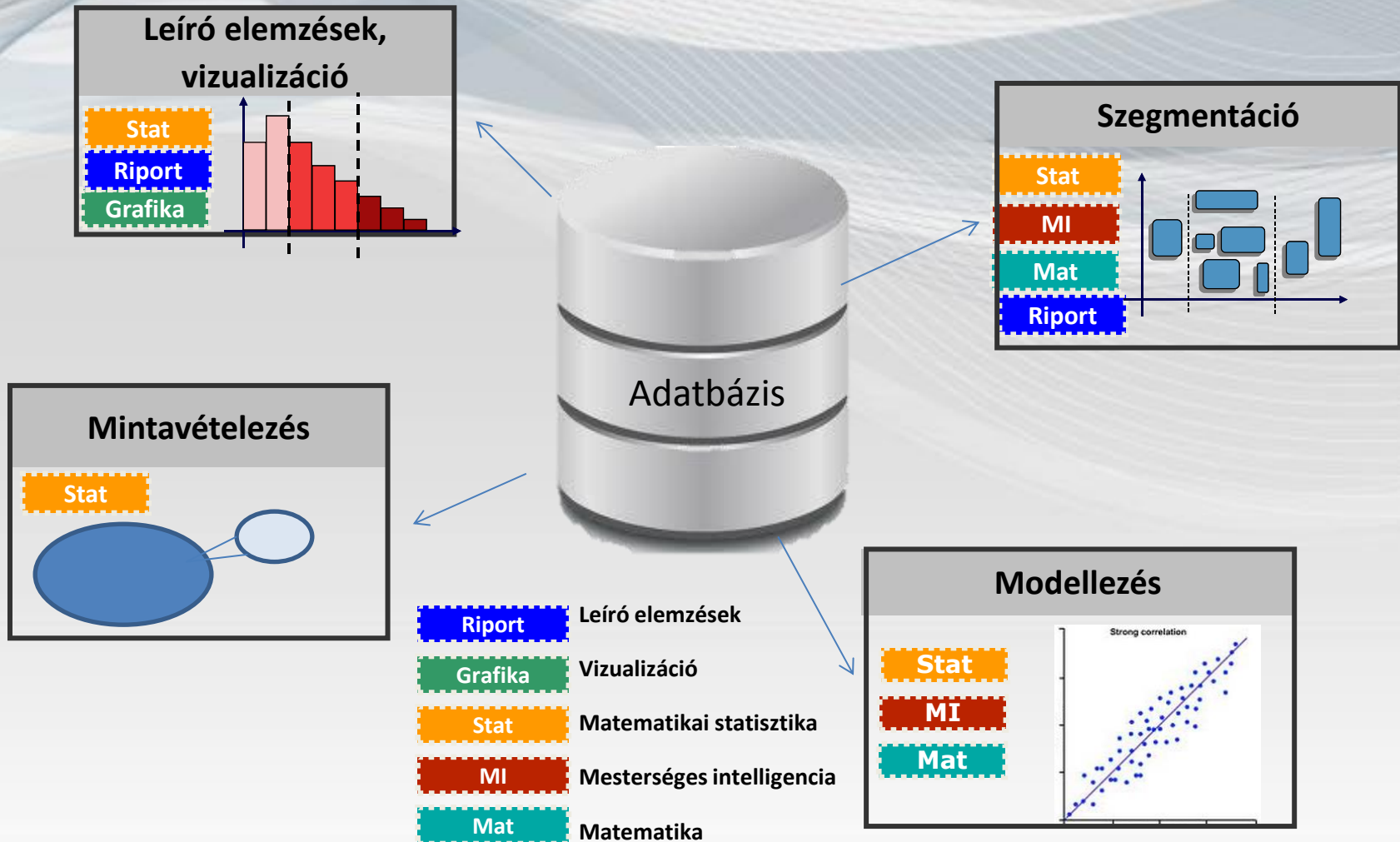
Az ügyes csaló feladata: „nem elég csalni, becsületesnek is kell látszani!”



itt látszani

itt lenni

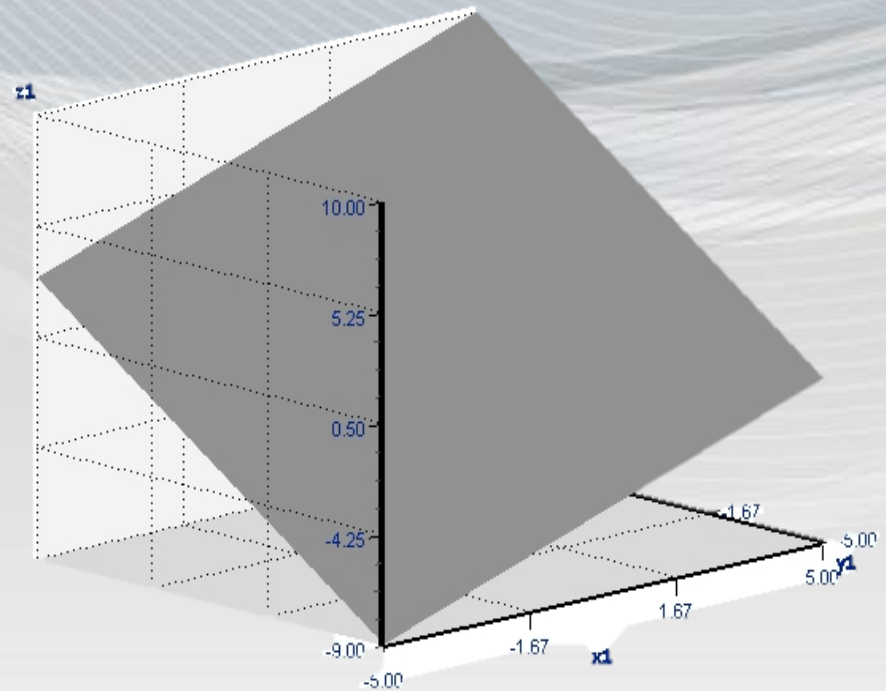
Az adatbányászat eszközszerze



Regresszió

REGRESSZIÓ:

- Magyarázó változók (**OK**) és az eredményváltozó (**OKOZAT**) közötti sztochasztikus kapcsolatot számszerűsíti
- Fundamentális okok feltárása
- Szigorú követelmény rendszer
- A modell formája: egyenlet
- Értelmezhető paraméterek, súlyok, score-ok



Logisztikus regresszió:

Modell kategóriás változóra,
Valószínűségek predikciója

Lineáris regresszió:

Folytonos változó predikciója

Megoldható a „nyomtalan csalás”?

Ha a csalásra létrehozott hálózatnak tetszőlegesen sok erőforrás áll a rendelkezésére, azaz **bármennyi megbízható tisztségviselő, alkalmazott, különböző székhely, telephely, bármennyi áru, és a csalárd tevékenység elfedésére alkalmas törvényes tevékenység áll rendelkezésére, és ismertek a hatóságok által alkalmazott elemzési módszerek,** akkor —megfelelő szakértelemmel— rövid ideig „sikeres” bűnszervezet működtethető. Ilyenkor minden ismert minta megjelenését elkerülhetik. A gépi tanulási módszerek is "nem deviáns" klaszterbe sorolják őket.

A feladatot mégsem oldották meg!

A rejtőzködés korlátja

- sok erőforrás:
 - veszteségbe fordítja a tevékenységet
- kevés erőforrás:
 - a) felismerhető minta (mintaillesztés)
 - b) kontrollált gépi tanulás
- a cél elárul:
 - költségvetés megkárosítása
 - haszon realizálása

ha mindenki
csal, akkor is
felismerhető

nem
kontrollált
gépi tanulás

önmagában
mintaképző

Mit tesz az informatika?

Attribútum vektor mindenből és mindenkiből

Súlyozások, normalizálások, linearizálás

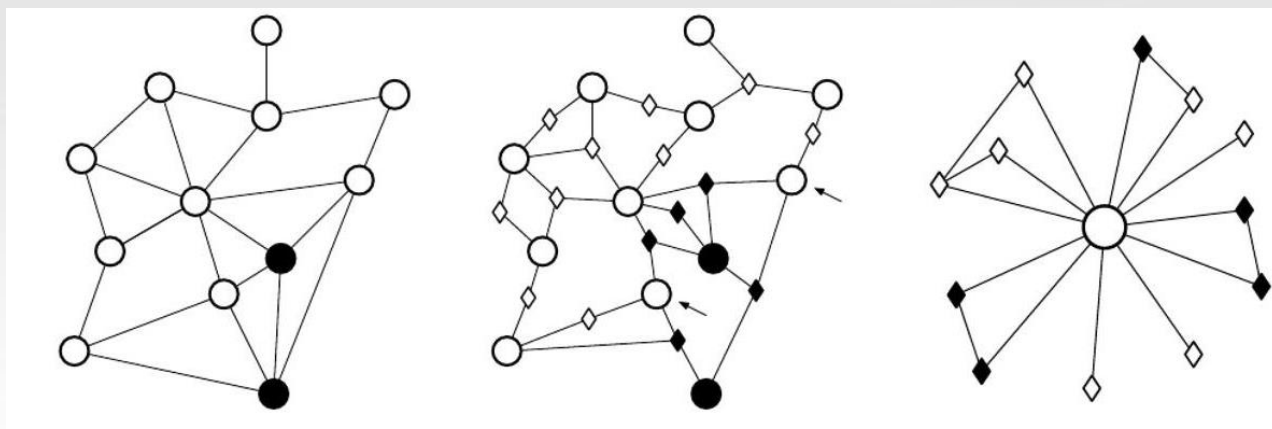
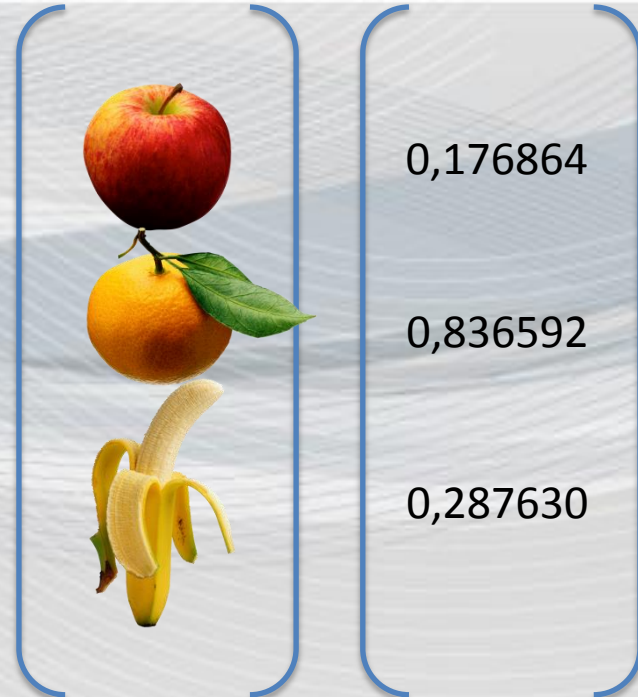
Összehasonlítások (pl. bezárt szög:
összeadjuk az almát a körével)

Kapcsolatok elemzése: család-terjedés
erőforrások elemzésével

Gráfminták
keresése, időbeni
változása

Machine learning

Kontroll !!!



Mi kell még?

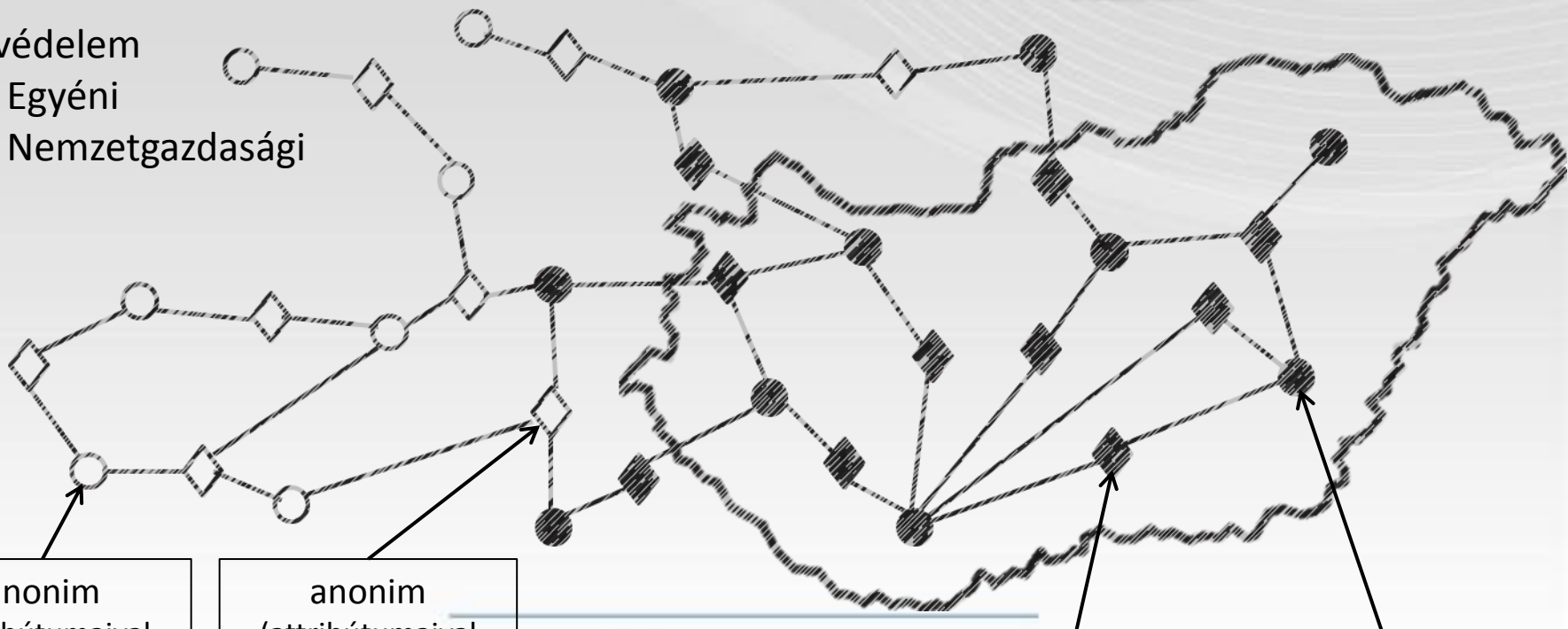


1. Még több adat!

2. Adat külföldről (legalább az EU-ból!)

Adatvédelem

1. Egyéni
2. Nemzetgazdasági



anonim
(attribútumaival
nem azonosítható)
gazdálkodó

anonim
(attribútumaival
nem azonosítható)
erőforrás

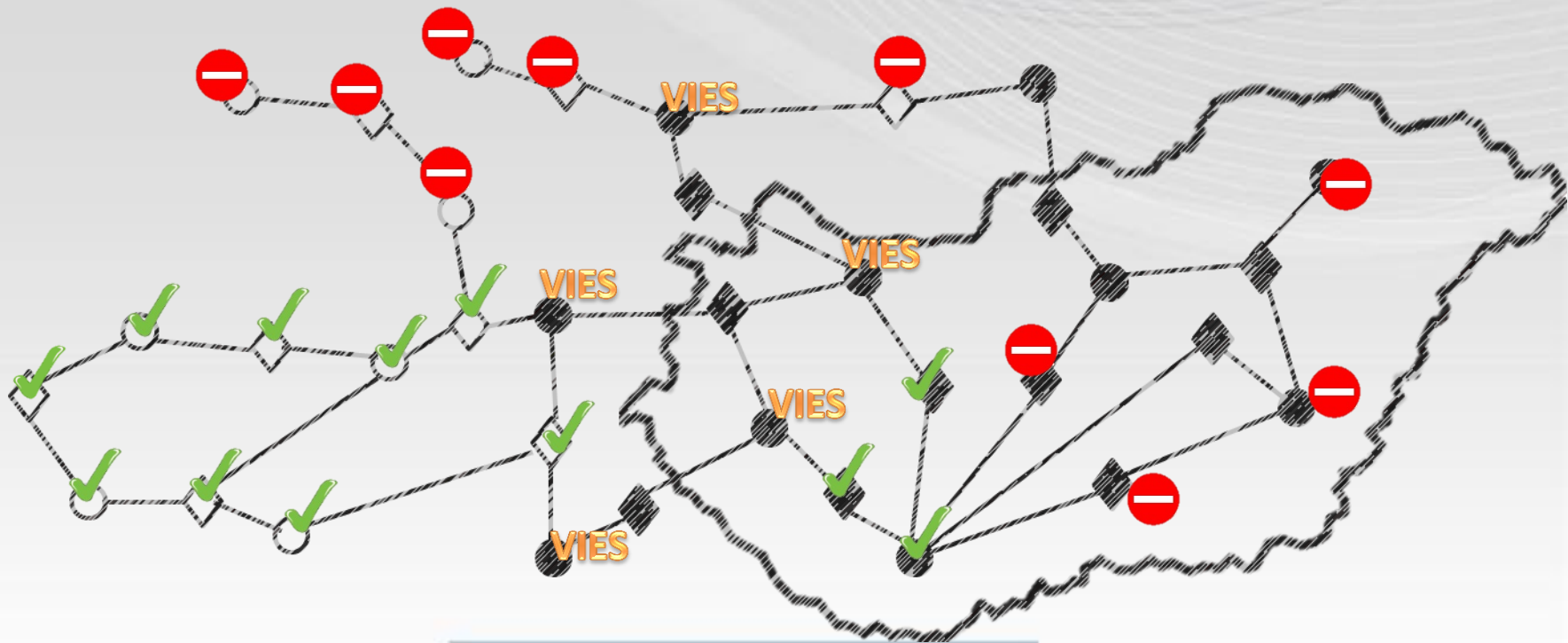


azonosítható
erőforrás

azonosítható
gazdálkodó

Szabályozás: fair anonimitás

- Nemzeti
- Közösségi anonimitás kölcsönös ESETI feloldása



Köszönöm a figyelmet



Nemzeti Adó-
és Vámhivatal