

VIRTUÁLIS KUTATÓ SZOBA PROJEKT, azaz **Big Data** módszertan medikai adathalmazokra

Klinikai betegéletút adatok mesterséges intelligencia
alapú kutatása projekt záró rendezvényre

Győr, 2020. január 29.

Király Gyula vezető kutató, *Hospitaly Kft.*

SZÉCHENYI 2020



MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Regionális
Fejlesztési Alap



BEFEKTETÉS A JÖVŐBE

Projekt azonosítója: **GINOP-2.1.1-15-2016-00898 K+F projekt**

Projekt időtartama: 2018.02.01 – 2019. 10.31

Projekt tárgya: „Orvos-szakmai és beteg életút adatok biztonságos kezelési protokollja, mesterséges intelligencia alapú orvosi minta vizsgáló szoba prototípus kifejlesztésén keresztül egy innovatív betegirányítási módszertan kialakítása”

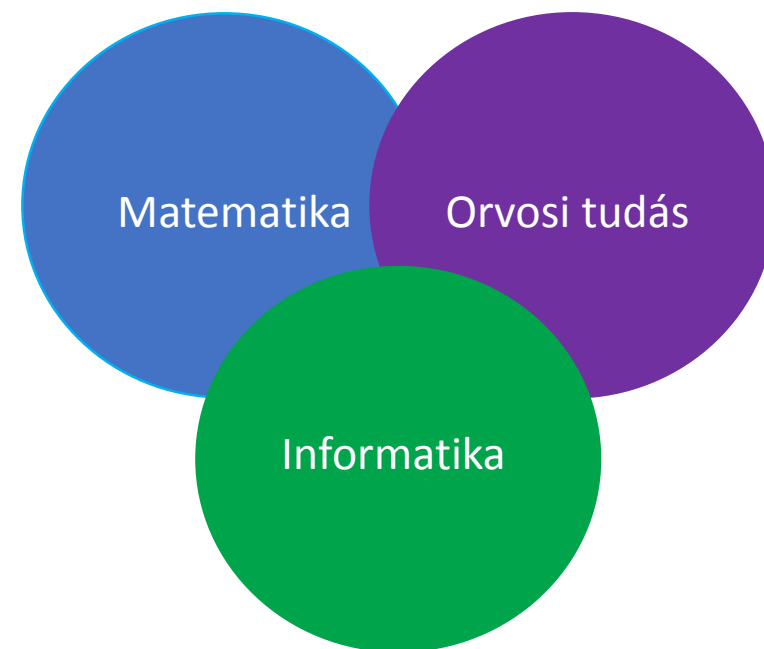
azaz

„Kutató szoba kialakítása, amelyben Big Data technológia felhasználásával egészségügyi intézményekben keletkezett és deperszonalizált egészségügyi adatok segítségével lehetőség van hipotézisek felállítására, bizonyítására vagy elvetésére.”

A projekt IT háttere: Szerver kapacitás - 440 mag, 5 T RAM, 10 db node, közösített tároló

- A kutatási feladat abból az alapvetésből indult ki, hogy egy innovatív, Big Data technológiára alapuló megoldást csak egy olyan szereplőkből álló kutató-team képes megvalósítani, amelyben egyenrangú és egymással magas színvonalon kommunikálni képes szakemberek találhatók.
- **Ezt a hármas egységet az orvos-szakmai, a matematikai és az informatikai szakterület képviselőiből állítottuk össze.**
- A projekt céljainak eléréséhez ezen felül szükség van jogi, közgazdasági valamint projektmenedzsment szakemberek aktív részvételére is.

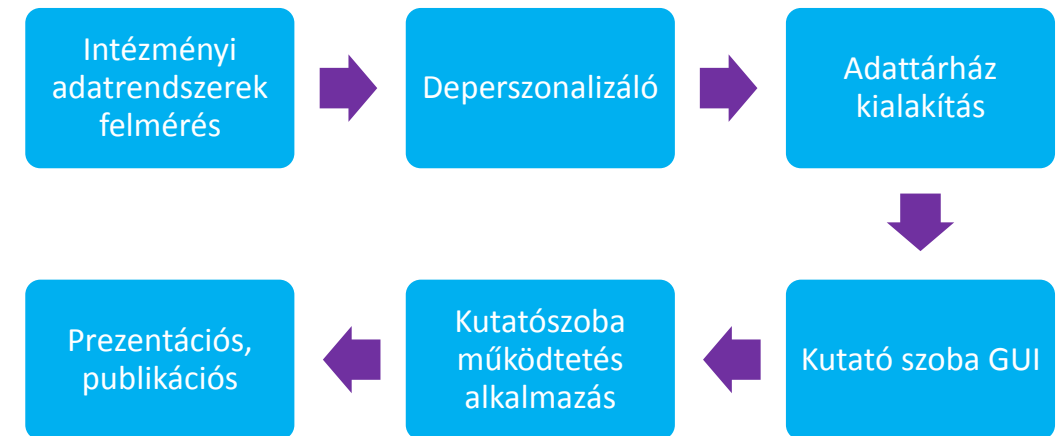
KUTATÁSI ALAPVETÉS



- A csapatban **nagy klinikai tapasztalatú orvosok, akadémiai kutató matematikusok** és országos egészségügyi rendszerek építésében **rutinos informatikusok** dolgoznak együtt különböző munkacsoportokban.
- Az adattárház kialakítása és üzemeltethetőség biztosítása az adattárház szakértők feladata.
- A projekt során kiemelten fontos az adatok védelme és az adatbiztonság, amit a jogi megfelelés munkacsoport biztosít.
- Az üzleti modell kidolgozását egészségügyi közgazdász irányítja.

PROJEKT SZERVEZET BEMUTATÁSA

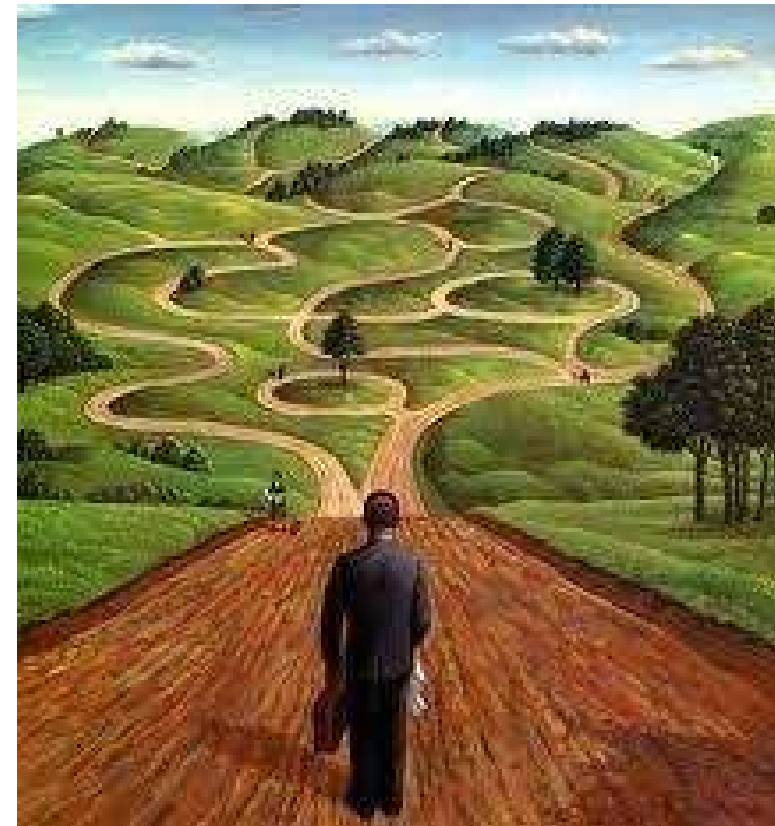
Vertikális munkacsoportok



Horizontális munkacsoportok



- Klinikai terület
 - Létező és nemzetközi szinten alátámasztott, de Magyarországon még nem megerősített klinikai hipotézis bizonyítása
 - Létező, de nagy számú mintán még nem igazolt klinikai hipotézis bizonyítása
 - Nem bizonyított klinikai hipotézis vizsgálata
- Közfinanszírozást érintő terület
 - HBCS súlyszámokat érintő vizsgálat
 - TVK korlátok hatásvizsgálata az ellátás minőségére
 - Közfinanszírozási gyakorlat hatásvizsgálata a várólisták hosszának növekedésére
- Minőségbiztosítási (belső működési) terület
 - Belső ellátási rutinok, protokollok vizsgálata
 - Méretgazdaságosságot érintő vizsgálatok
 - Szakmai kompetencia vizsgálatok
 - Terápiák, beavatkozások, vizsgálatok hatékonyságának ellenőrzése
 - Szakmai protokollok ellenőrzése, megerősítése vagy elvetése
- Népegészségügyi terület
 - Életmódhoz köthető betegségek vizsgálata
 - Életmód által befolyásolt terápiák hatásosságának vizsgálata
 - Populációt érintő kockázati faktorok vizsgálata
- Egyéb
 - Informatikai megoldás és célzott adatgyűjtés torzító hatása
 - Adatvalidálás fejlesztésének lehetőségei



Egészségügyi adatok körének meghatározása

- A projekt végcéljának, azaz egy kutató szobának a megvalósításához a megfelelő infrastruktúra biztosítása mellett olyan mennyiségű és minőségű adatokra van szükség, amelyek kielégítik az orvosszakma által felállított hipotézisek alátámasztásának vagy elvetésének feltételeit.
- Egy adott egészségügyi intézmény klinikai informatikai rendszerének adatbázis struktúrájának felmérése után kiválasztottuk azokat az adatokat, amelyek egyértelmű értelmezéséhez rendelkezünk megfelelő eszközökkel (kódok, dátumok, numerikus értékek, egyéb klasszifikált adatok, stb.)



Személyes adatok kiváltása demográfiai adatokkal

- Az intézményen belül a megfelelő infrastruktúrán zajlik a TAJ szám és a hozzá tartozó személyes adatok leválasztása.
- Eközben megtörténik az eset azonosítónak a kutatáshoz szükséges egyes demográfiai jellemzőkkel történő felruházása és kiegészítése, közvetlenül vagy adat-transzformációval.
- A kutatás kiterjed a releváns anamnézis adatok esetleges bevonásának lehetőségére is.

PROJEKT FONTOSABB MÉRFÖLDKÖVEI

TAJ szám leváltása életút azonosítóval

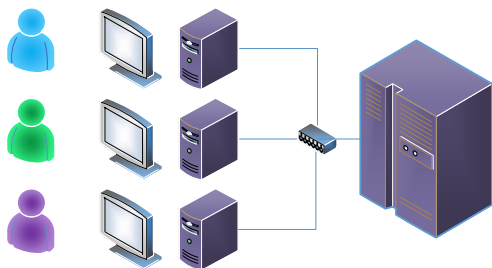
- A kutatás során kialakításra került egy intézmény független, biztonságos, irreverzibilis betegéletút azonosító képzési eljárás és hozzá tartozó módszertan.
- Még intézményen belül megtörténik a TAJ szám átalakítása betegéletút azonosítóvá a módszertanban definiált zárt, többfaktoros, kriptográfiai eljárás segítségével.



HASHCODE

Kutatási adattárház kialakítása

- Az összegyűjtött és betegéletút azonosításra alkalmassá tett, deperszonalizált adathalmaz átkerül a kutatási adattárházba.
- Elkészültek az orvos-szakmai igényeknek megfelelő adatpiacok, előkészítésre került a matematikusok által megadott függvények és eljárások futtatásához szükséges környezet.
- Kidolgozásra került a folyamatos adatbetöltés módszertana.



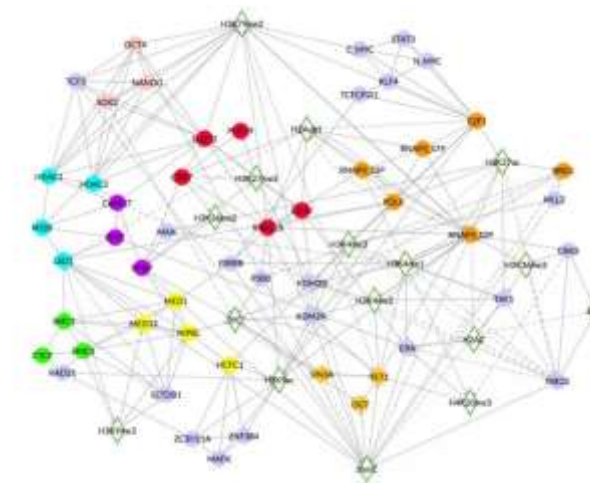
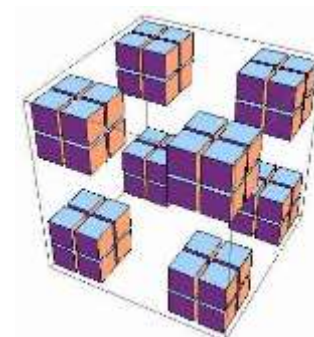
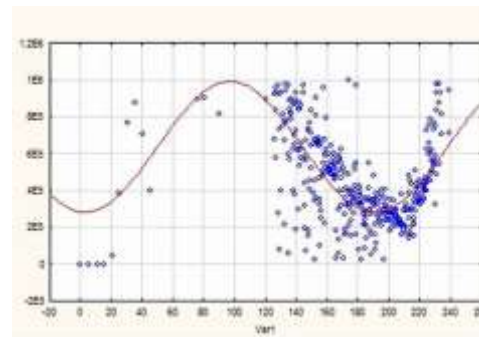
PROJEKT FONTOSABB MÉRFÖLDKÖVEI

Működési modell, publikációs rend kidolgozása

- Intézményenkénti adatvagyon nyilvántartás – kutatási algoritmus által felhasznált adatkör szerinti elszámolás.
- Folyamatos csatlakozási lehetőség, ország- és nyelvfüggetlenség, zárt kutatási adminisztráció és elszámolás.
- Tudományos megállapítás egyéni vagy intézményi hatáskör, publikálás közös eredmény



- A projekt eredményeként, illetve a bemutatott módszertan segítségével lehetőség nyílik arra, hogy nagy mennyiségű adathalmazokon teszteljünk orvosi hipotéziseket valamint eddig nem vizsgált, de gyanított összefüggéseket támasszunk alá a modern statisztika és adatbányászat olyan eszközeivel, mint pl. a support vector machine, főkomponens elemzés (Principal Component Analysis - PCA) és egyéb gépi tanulós algoritmusok.
- Ezáltal az orvosi praktikumban egyszerűbb elemzések alapján is megbízhatóbb előrejelzést lehet adni, csökkentve az extra negatív vizsgálatok számát és gyorsítva a diagnózisok megállapítását.



Adattárház előkészítés:

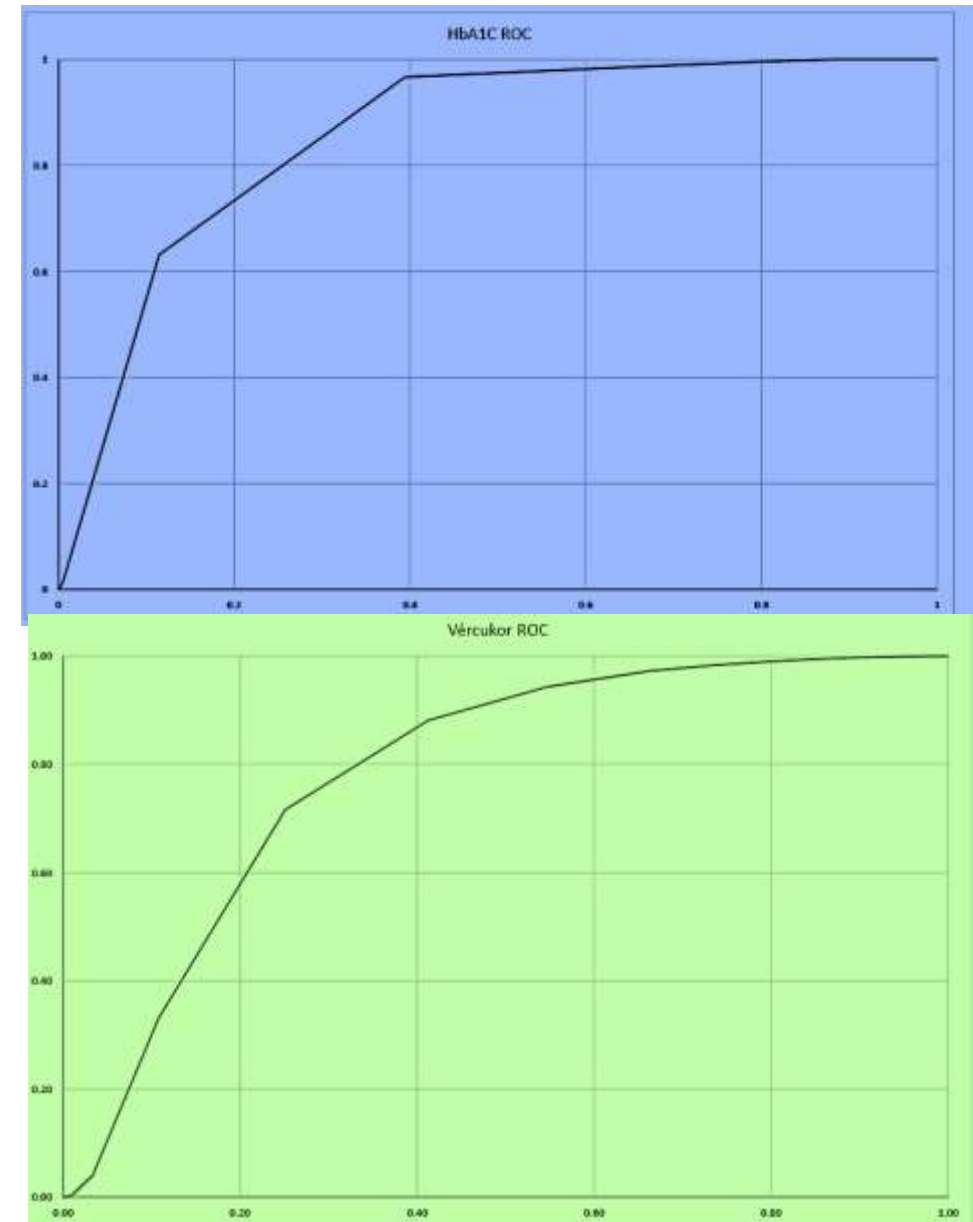
- ~600 ezer beteg, ~17 millió eset, ~44 millió releváns laborvizsgálati eredmény leszűrése

Matematikus felvetés:

- A HbA1C értékek alapján jobban el lehet szeparálni a cukorbetegeket a nem cukorbetegektől
- AUC érték 0.86, míg a vércukor alapján az AUC 0.79

Matematikus kérdés és a **Klinikus válasz:**

- Lehet-e a relatíve alacsony AUC érték oka, hogy a kezelt cukorbetegek között néhánynak a vércukorszintje normális értékű? – **Igen!**
- Lehet-e, hogy cukorbetegséggel nem diagnosztizált páciensek egy része valójában cukorbeteg? – **Igen!**

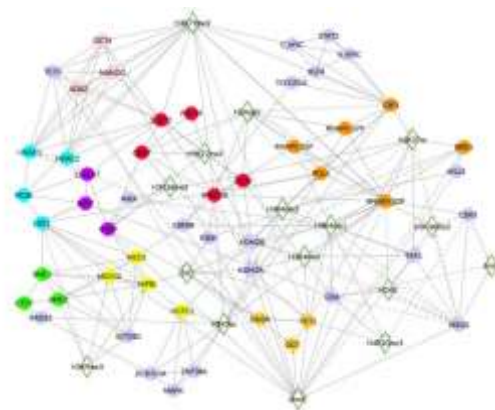
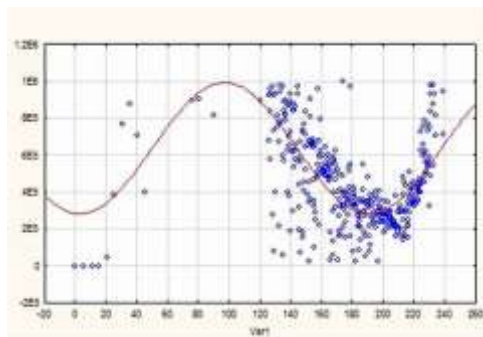


Klasszifikált, idősoros egészségügyi adatok valamennyi magyarországi intézményben

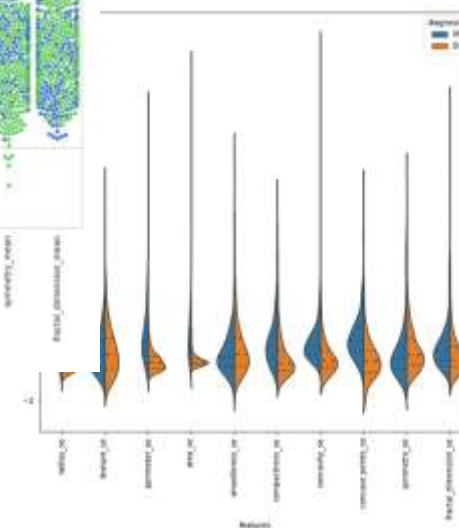
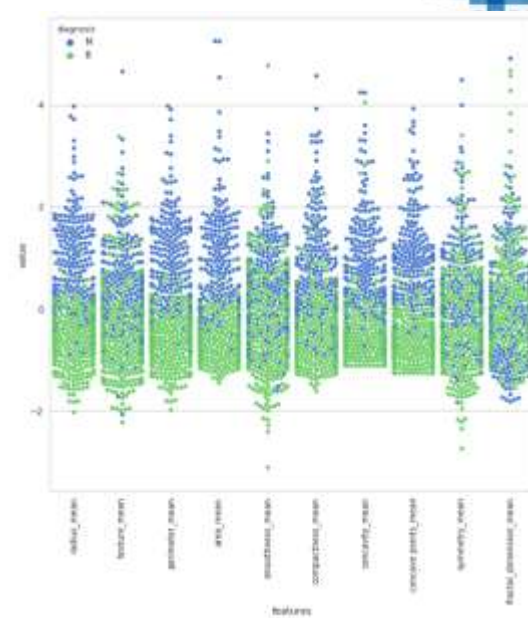
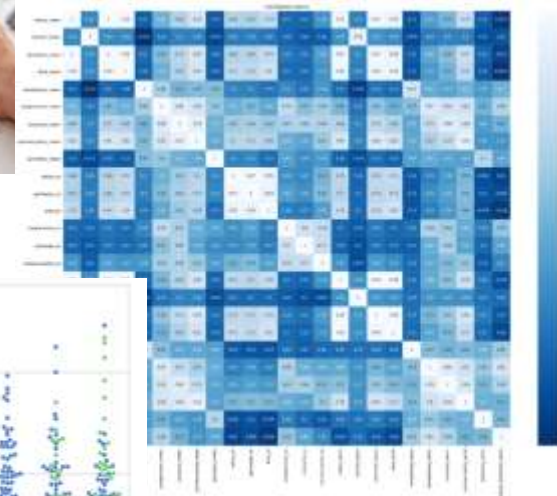
- Létező országos beteg azonosító rendszer (TAJ szám)
- Nagy mennyiségű és egységes minőségű adatok valamennyi intézményi medikai rendszerben (HIS)
- Matematikailag jól értelmezhető natív adatok (kódok, dátumok, numerikus értékek, egyéb klasszifikált adatok, stb.)
- Jól körülhatárolható demográfia adatok minden betegnél
- Országosan egységes közös halmaz a tételes finanszírozási jelentés



- Orvosegyetemi képzés kibővítése a matematika, matematikai statisztika és a BigData ismeretekkel
- Közös work-shop-ok szervezése, ahol megismerhetik egymás gondolkodásmódját, elvárásait
- Adat vizualizáció fejlesztése - a modern statisztika és adatbányászat riportjainak megismeréséhez (pl. a support vector machine, főkomponens elemzés (Principal Component Analysis - PCA) és egyéb gépi tanulós algoritmusok)



LEHETŐSÉGEK



Problémák, kérdések?

- Mit érnek az adatok betegségút kapcsolat nélkül?
- Ki gyűjthet adatot valamennyi intézményből?
- Meg lehet-e szerezni valamennyi beteg hozzájárulását?

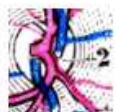
Lehetséges megoldások

- Törvényi felhatalmazás
- TAJ szám átalakítása betegségút azonosítóvá
- Retrospektív kutatási adatbázis építése
- Egyszeri adatbázis mintavétel és TAJ mentesítés (sokszor)



Nyílt adatbázisok – Közösségi kutatás

- A legnagyobb pénzdíjas befejezett orvosi



Diabetic Retinopathy Detection

Identify signs of diabetic retinopathy in eye images

Featured · 4 years ago · optometry, health sciences, image data, binary classification

\$100,000
661 teams

- A legtöbb kutatói csoport által letöltött



Titanic: Machine Learning from Disaster

Start here! Predict survival on the Titanic and get familiar with ML basics

Getting Started · Ongoing · tutorial, tabular data, binary classification

Knowledge
11,413 teams

- A legnagyobb pénzdíjas nyitott kutatás

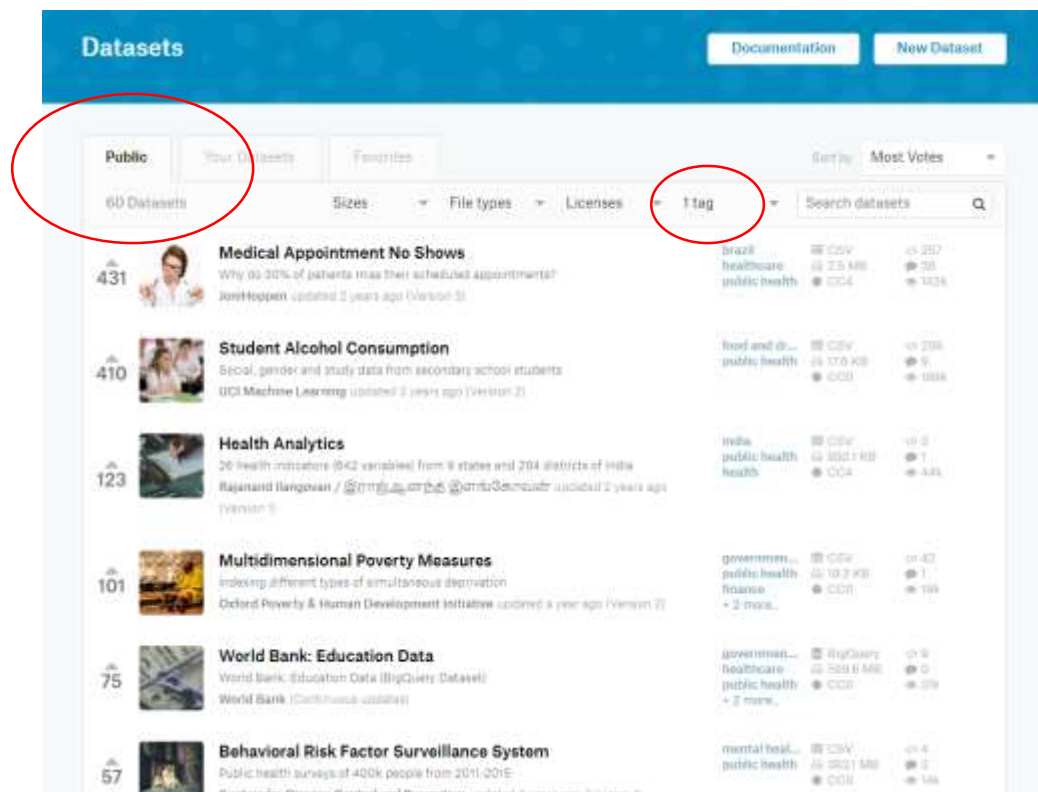


LANL Earthquake Prediction

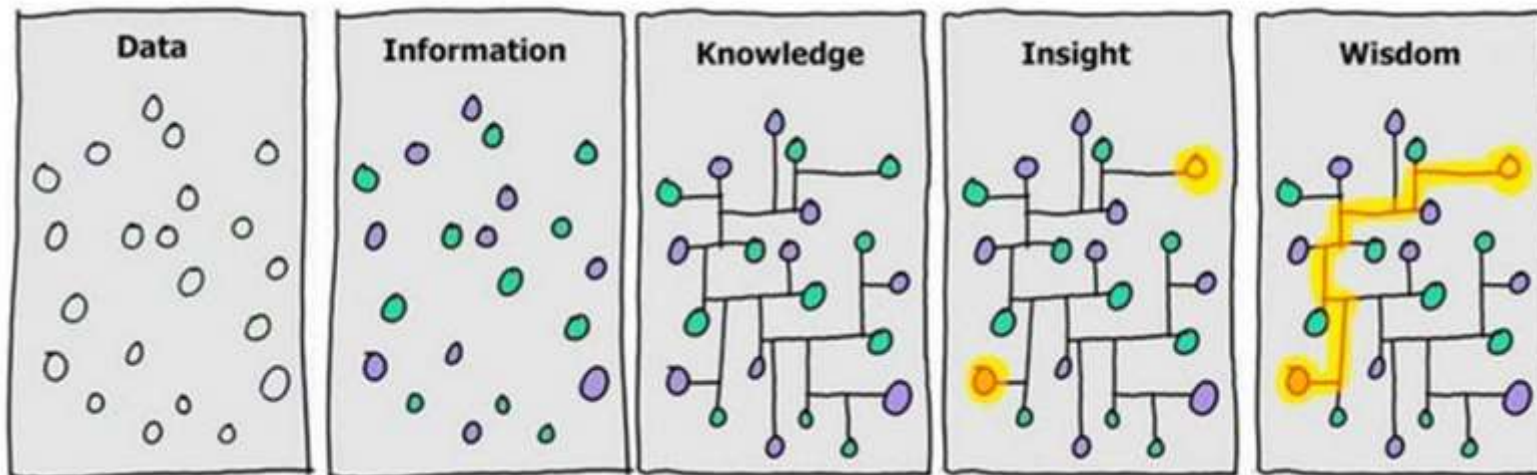
Can you predict upcoming laboratory earthquakes?

Research · 17 days to go · earth sciences, physics, signal processing

\$50,000
4,009 teams



KÖSZÖNÖM A FIGYELMET!



SZÉCHENYI 2020



MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Regionális
Fejlesztési Alap



BEFEKTETÉS A JÖVŐBE