

Hogyan lesz adatbányából aranybánya?

Szolgáltatások kapacitástervezése
a Budapest Banknál

Németh Balázs – Budapest Bank
Fehér Péter – Corvinno
Visontai Balázs – KFKI

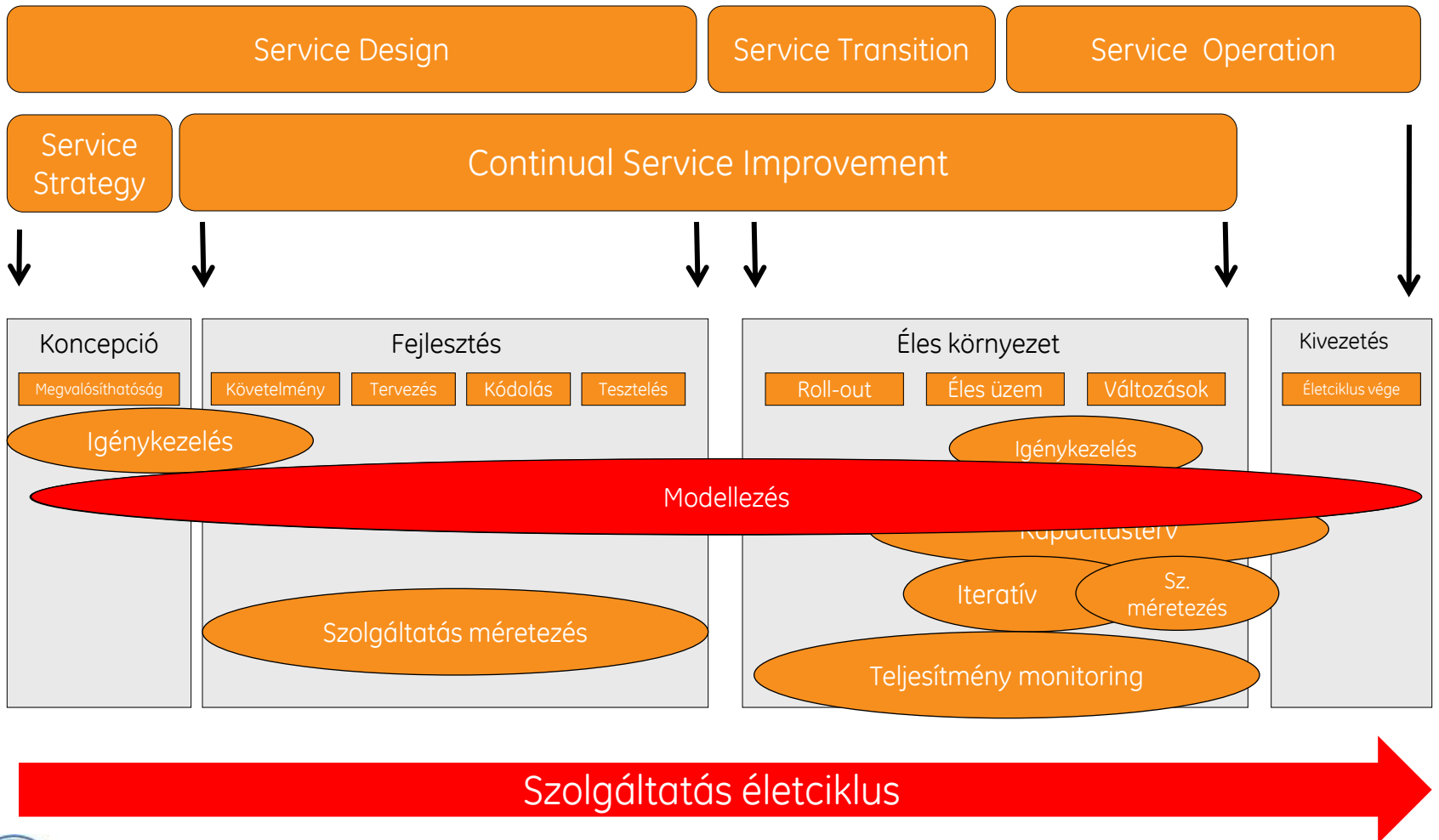


Tartalom

1. Szolgáltatás életciklus
2. Előzmények
 1. Kihívások
 2. Fókuszterület
3. A projekt bemutatása
4. Tapasztalatok



Szolgáltatás életciklus

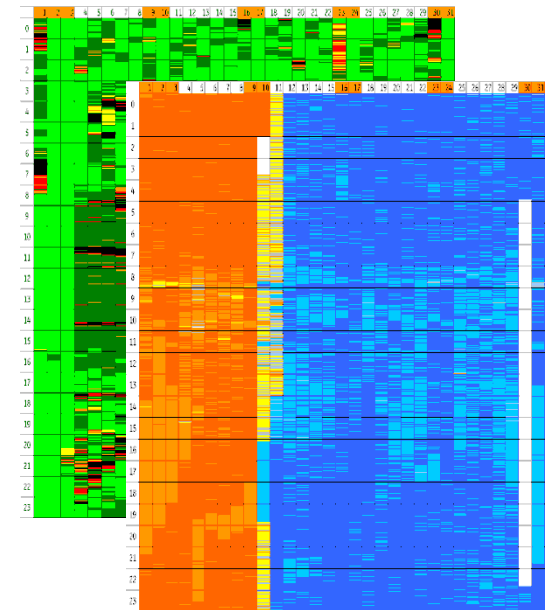


ELŐZMÉNYEK

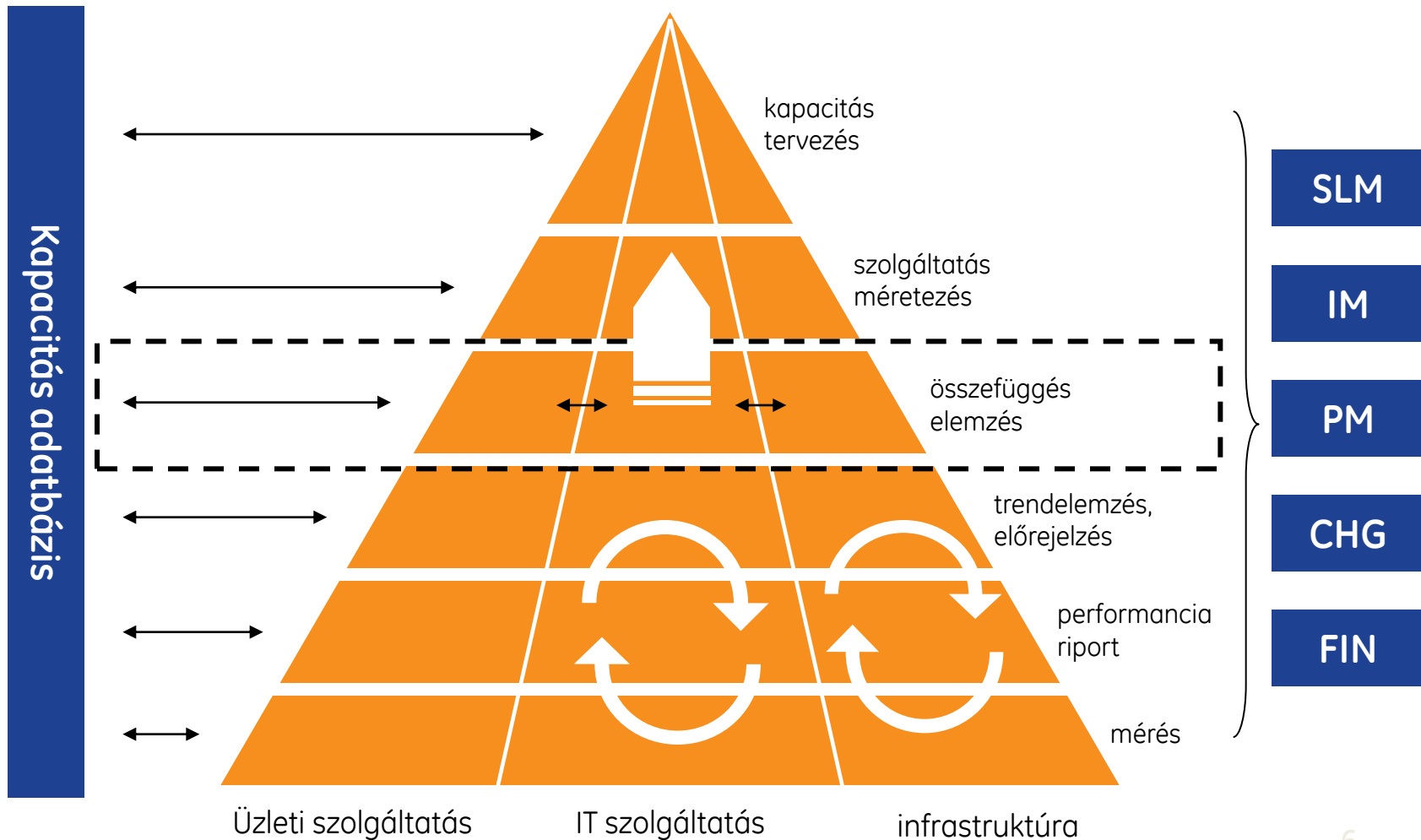


Kihívások kapacitásmenedzsment területen

- Komplex és heterogén architektúra
- Osztott erőforrások
- Feldolgozott és riportált adatmennyiség
 - Infrastruktúra performancia
 - Szolgáltatás performancia
- Feldolgozott és riportált adatkörök egymásra hatása nem ismert

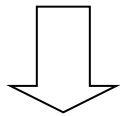


Kapacitásmenedzsment funkciók

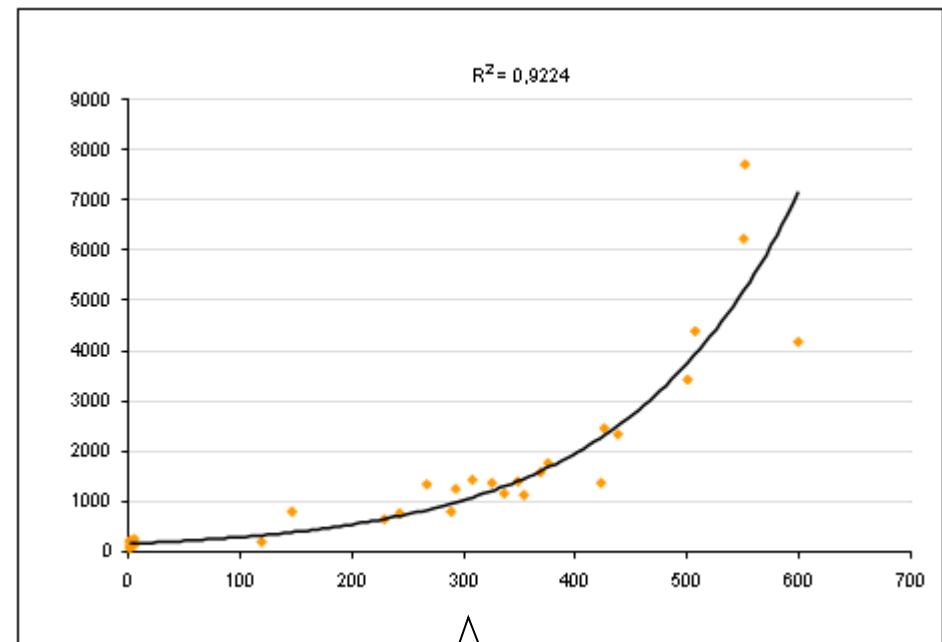


Összefüggés elemzés – statisztikai alapokon

- Trendelemzés, előrejelzés
- Összefüggés elemzés



- Erőforrás becslés
- Performancia képesség meghatározás
- SLA javaslatok
- Szűk keresztmetszetek megállapítása



Magas szintű
összefüggés volumen és
terhelés között

A PROJEKT



Aktuális kihívások

- Számos mérés történik
- Nagy mennyiségű adat keletkezik
- Adatok felhasználása
 - Döntéshozatal
 - Beavatkozási lehetőségek
 - Fejlesztési tervek
- A szolgáltatások viselkedése terhelések változására nem ismert
- Ráfordítások és szolgáltatási szintek között nincs kimutatható kapcsolat



Kutatás részei

Adatkörök vizsgálata

- Jelenlegi környezet vizsgálata
- Operációs rendszer, tranzakciók, alkalmazási adatok

Mutatószámok vizsgálata

- Mutatószámok összegyűjtése és vizsgálata
- További mutatószámok használatának vizsgálata

Támogató rendszerek vizsgálata

- Modellek alkalmazásához szükséges támogató eszközök

Összefüggések elemzése

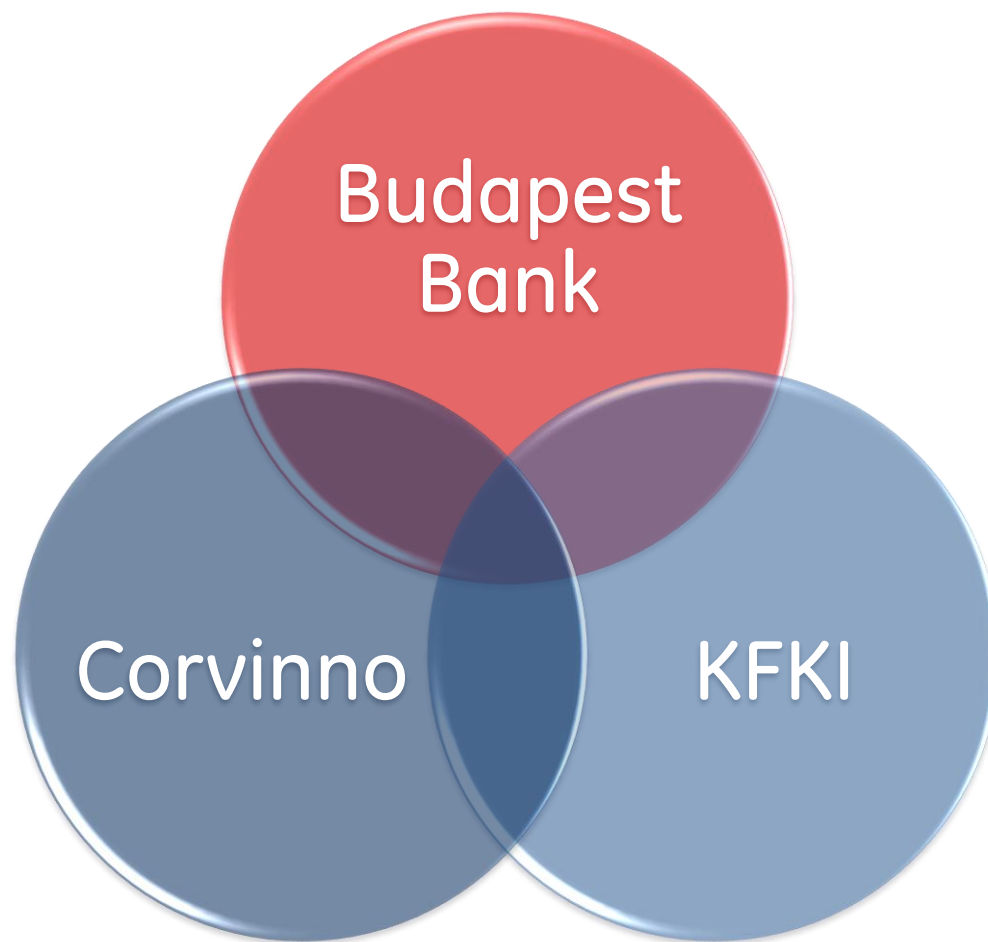
- Modellezési megoldások vizsgálata
- Modellezés végrehajtása

Forgatókönyv elemzési modellek

- Trendek, események hatásainak vizsgálata



Projekt felépítése



Vizsgálati pilot terület

- Szignifikáns üzleti hasznosság
- Vizsgált terület komplexitása
- Vizsgált területhez kapcsolódó infrastruktúra feltárása
- Kialakított mérések
 - Széles körű lefedés
 - Nagy adatmennyiség
- SLA álljon rendelkezésre



Vizsgált adatkörök

Operációs rendszer szintű adatkörök

Processzor leterheltség
Memória leterheltség

Adatbázis szerver szintű adatkörök

Diszk válaszidő
Várakozási sorok
Írás, olvasás idő

Tranzakció szintű adatkörök

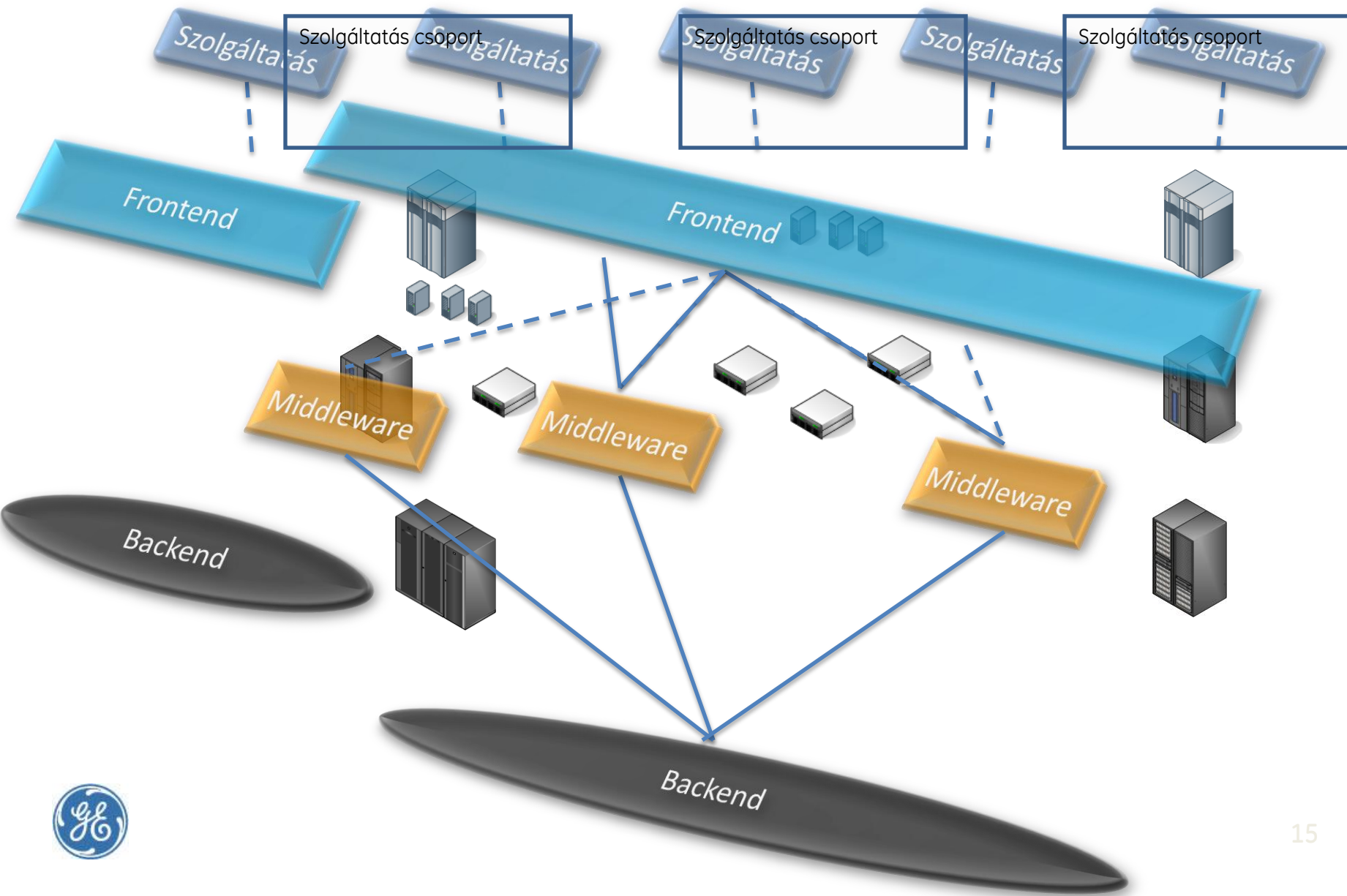
Tranzakciók típusai
Tranzakciók száma
Felhasználók száma
Tranzakció válaszidők

Hálózati szintű adatkörök

Hálózati eszközökön átmenő forgalom
Hálózati eszközök processzor terheltsége
Hálózati eszközök memória terheltsége



Támogató architektúra



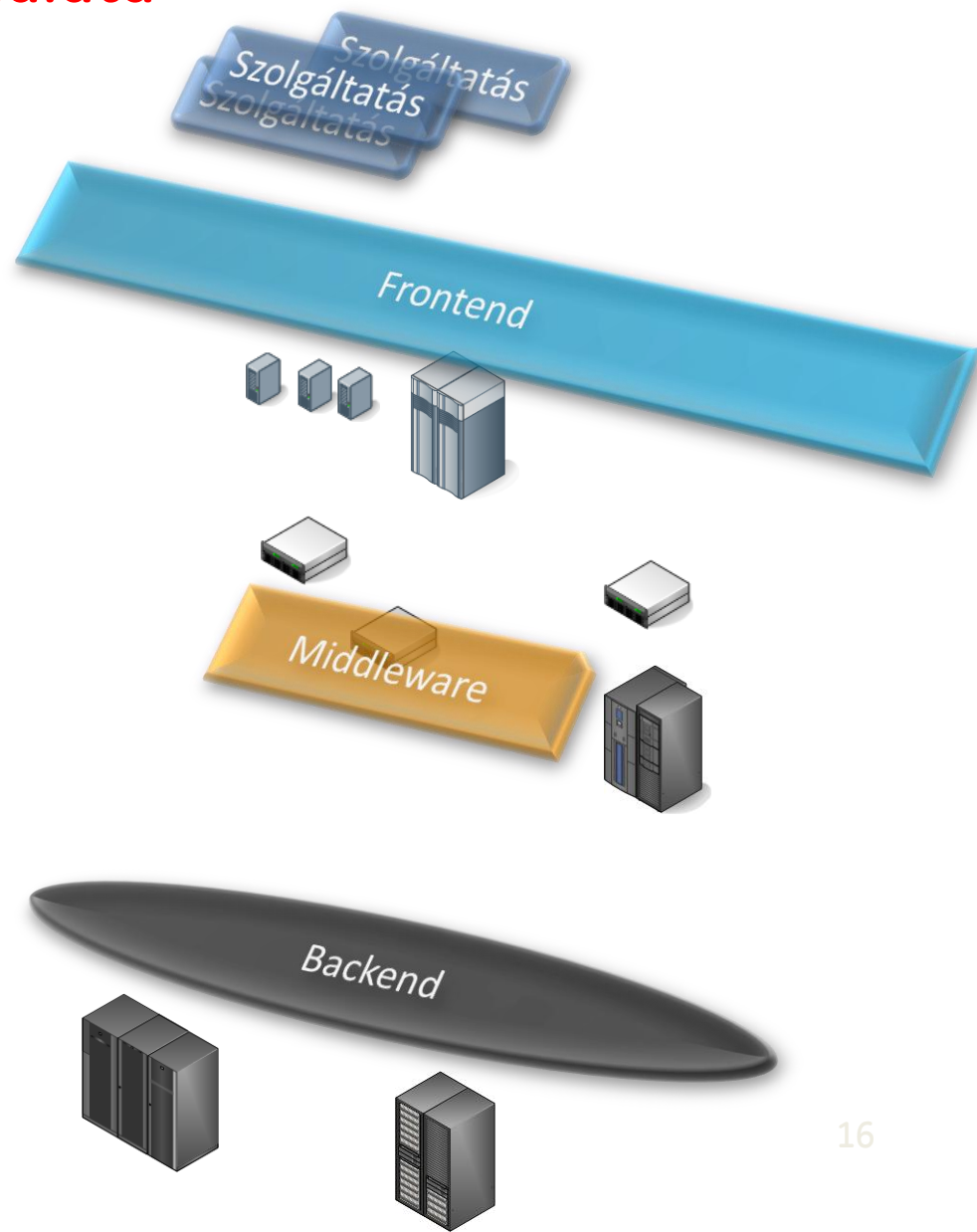
Szolgáltatáscsoportok vizsgálata

Tranzakció szintű adatkörök

Tranzakciók típusai
Tranzakciók száma
Tranzakció válaszidők

Operációs rendszer szintű adatkörök

Processzor leterheltség



Vizsgált módszerek, eljárások

- Alapstatisztika
 - Korreláció
 - Regresszió
- Sokváltozós statisztika
 - Faktorelemzés
 - Klaszterelemzés
- Neurális hálók alkalmazása



Eszközök, adatok



Statisztikai
összefüggések

Modellalkotás,
modellfuttatás,
előrejelzés



2 hónapnyi adat



Az elemzés végrehajtása

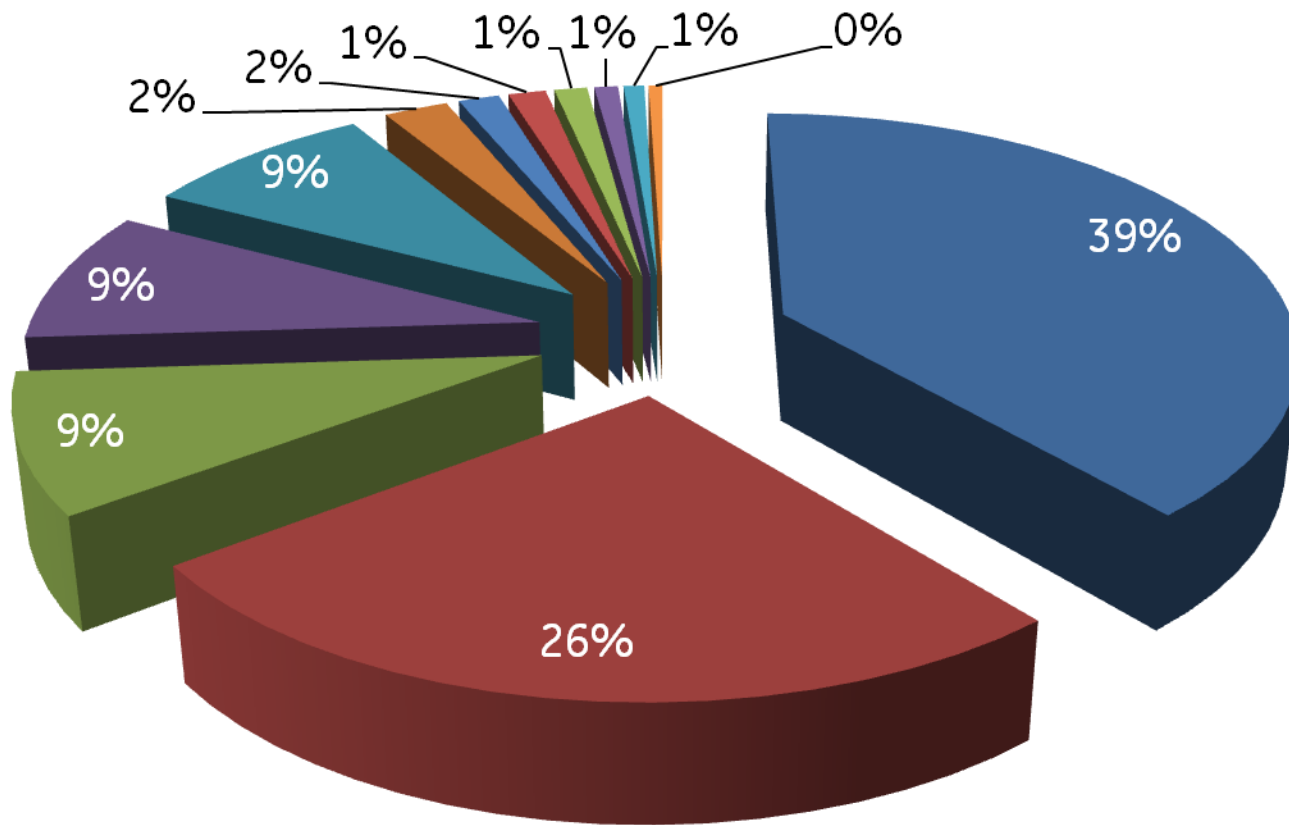
- Pilot terület meghatározása
- Tranzakciók arányának vizsgálata
- Rendelkezésre álló mérések vizsgálata
- Adatok megfelelősége
- Elemzési lehetőségek vizsgálata
- Elemzések végrehajtása, modellépítés
- Eredmények tesztelése



ÁLTALÁNOS EREDMÉNYEK



Vizsgált tranzakciótípusok megoszlása

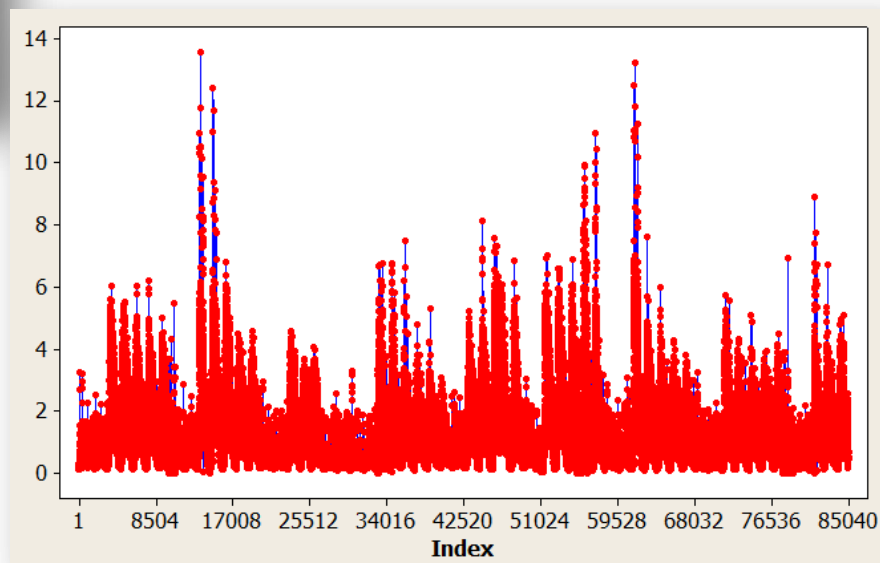
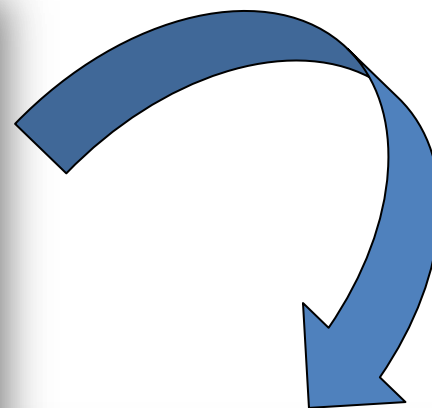
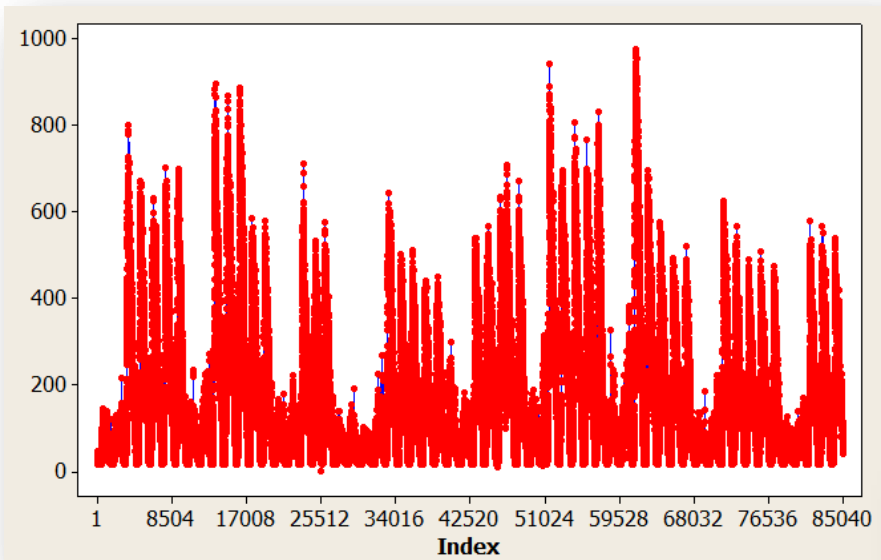


Szolgáltatásokhoz kapcsolódó elemzések elvárásai

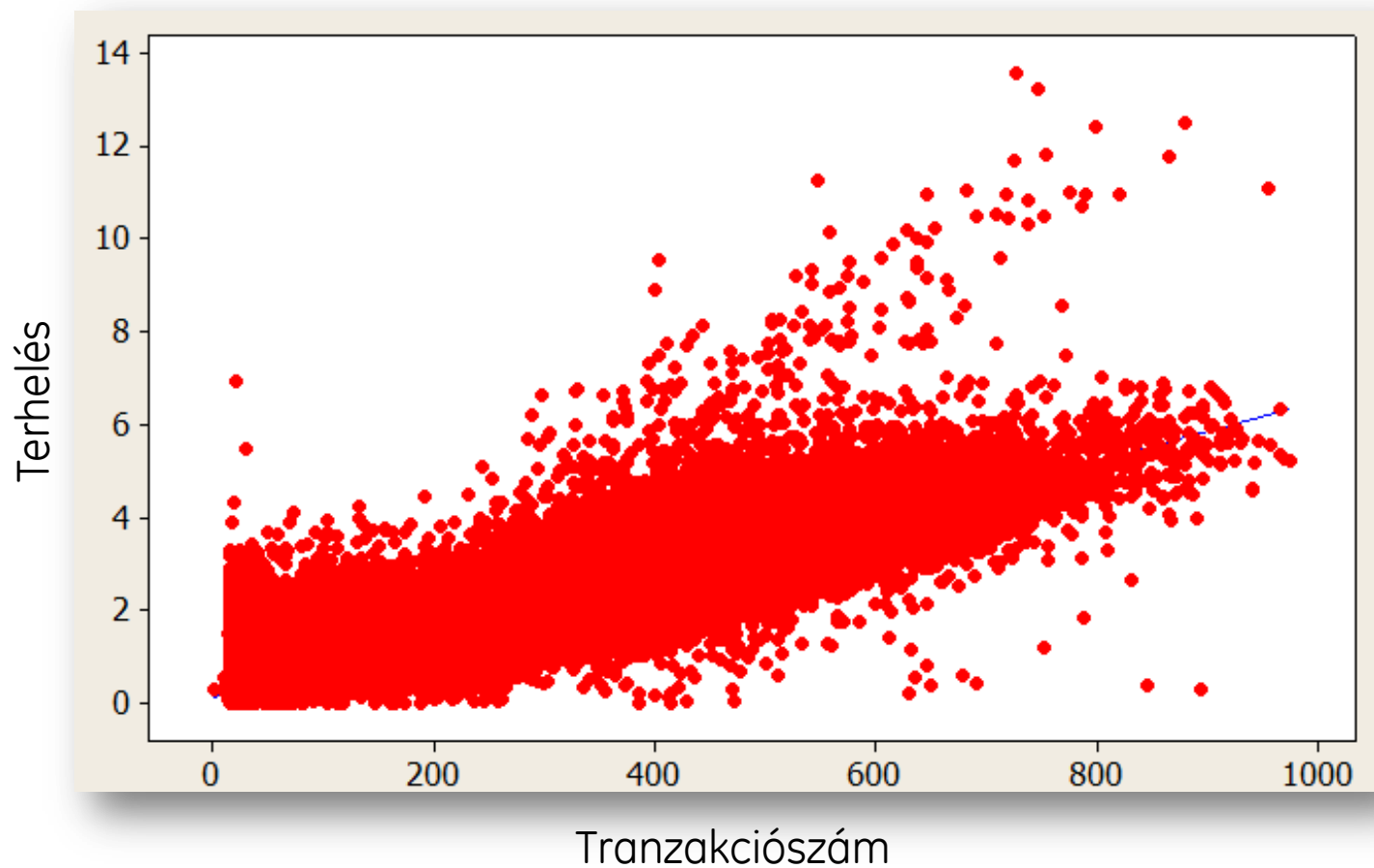
- Mért adatmennyiség
- Adatok reprezentativitása
- Terhelt infrastruktúraelemek, feldolgozási út



Tranzakciószám és front-end terhelés



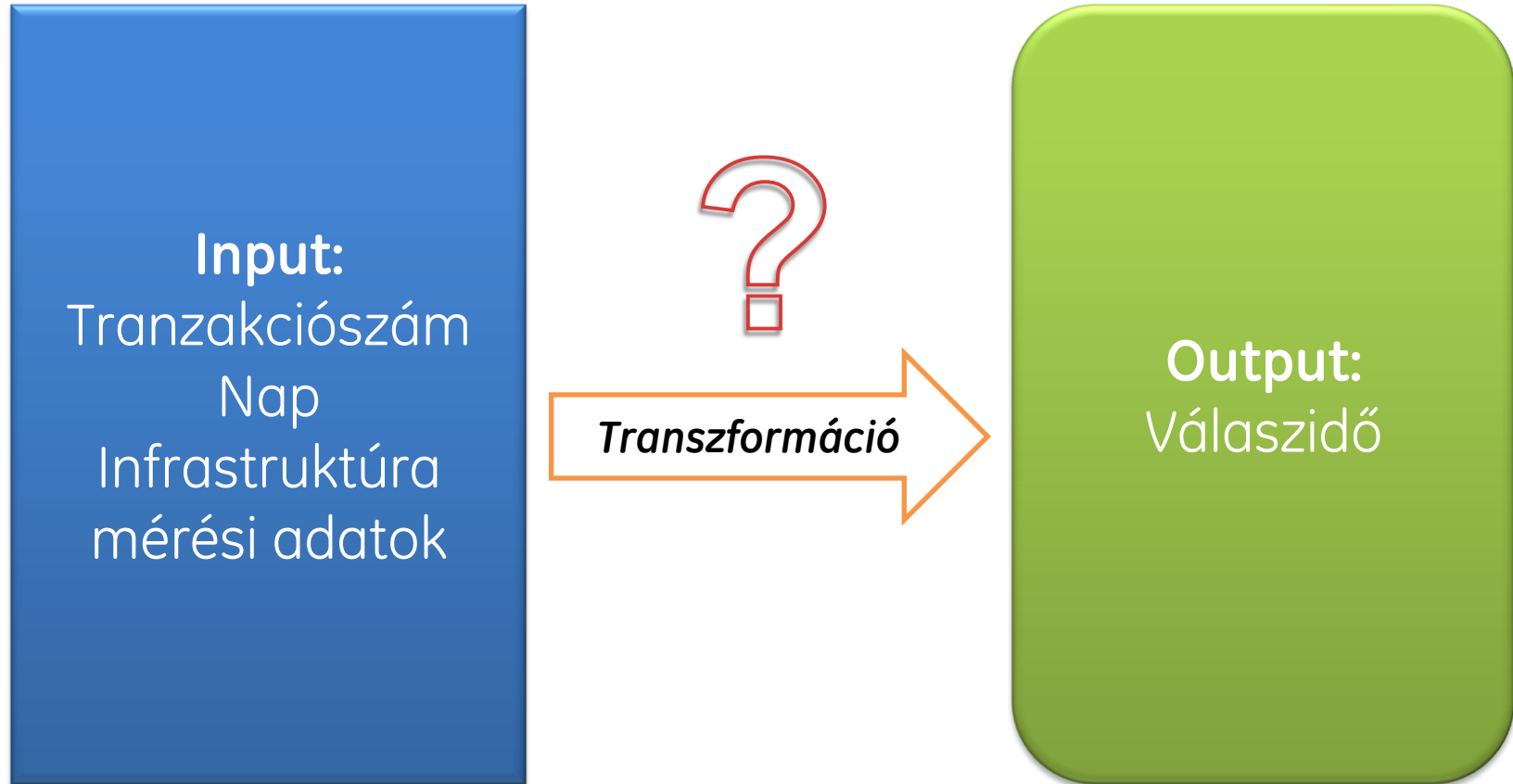
Összefüggések statisztikai vizsgálata



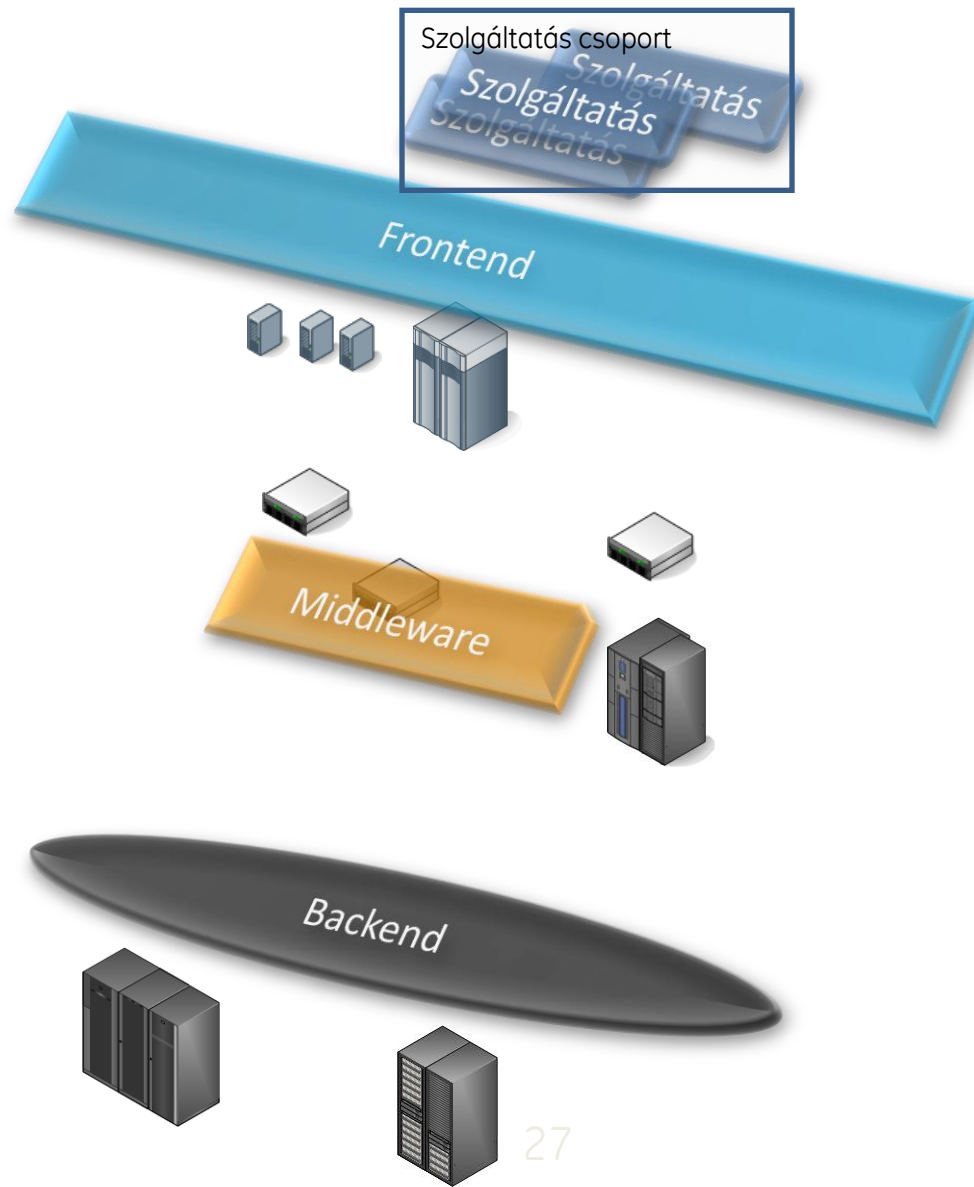
AZ ALAPSTATISZTIKÁK KORLÁTAI



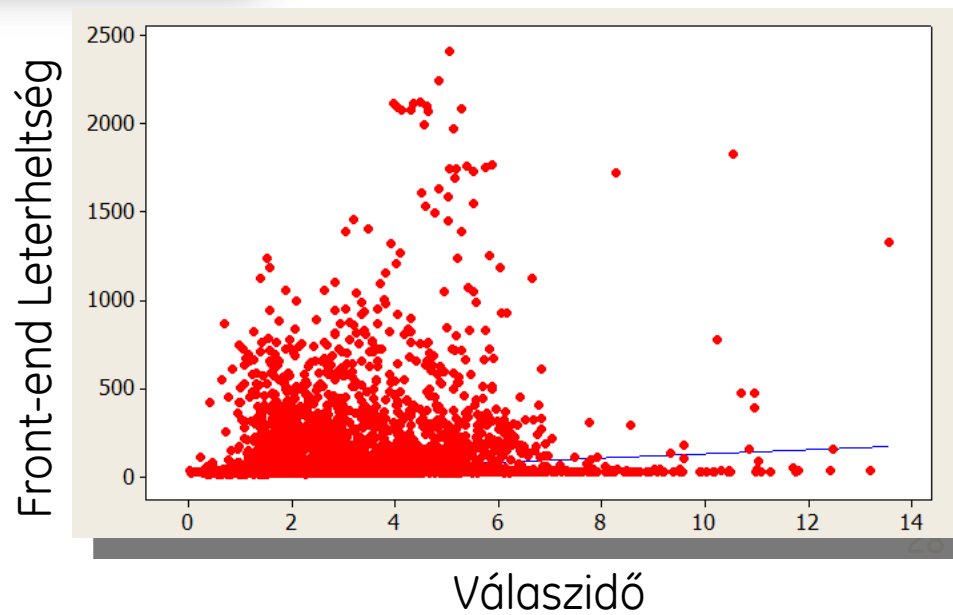
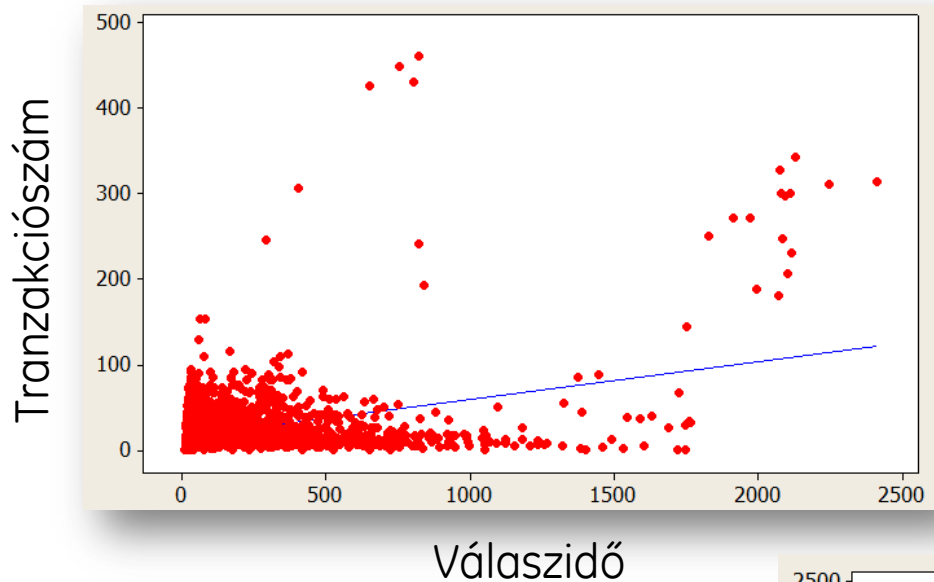
Modell kialakítás



Szolgáltatáscsoportok vizsgálata



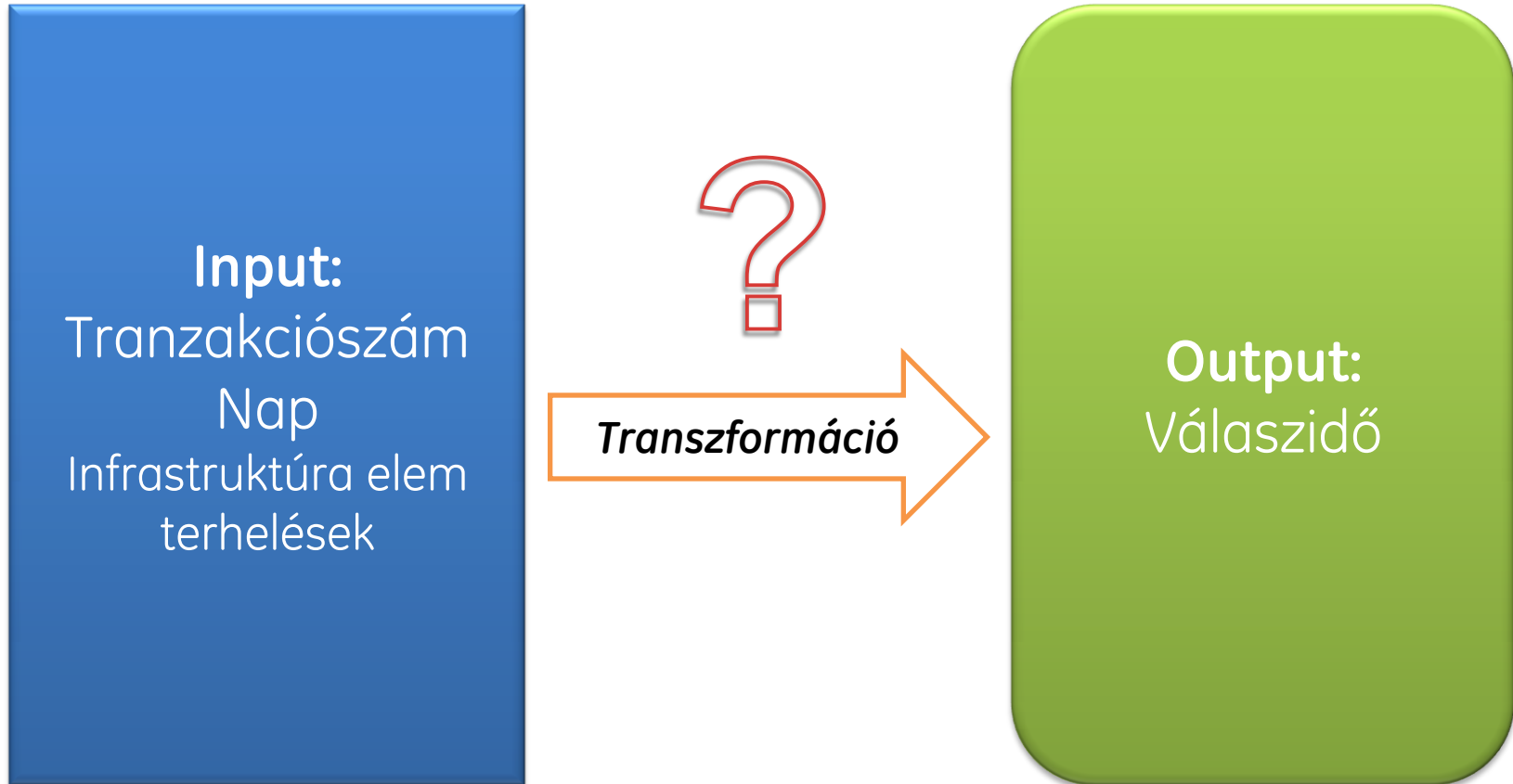
Összefüggések



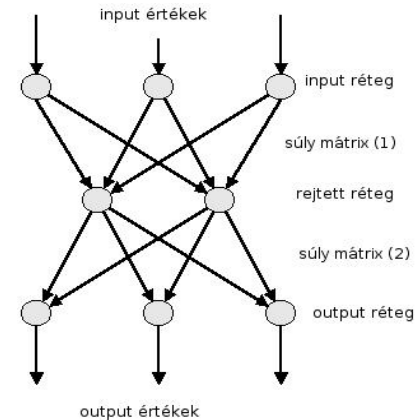
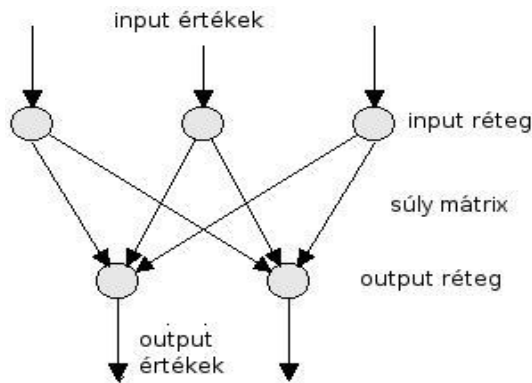
MODELLEZÉS EREDMÉNYEK



Neurális háló leképezés



Neurális háló

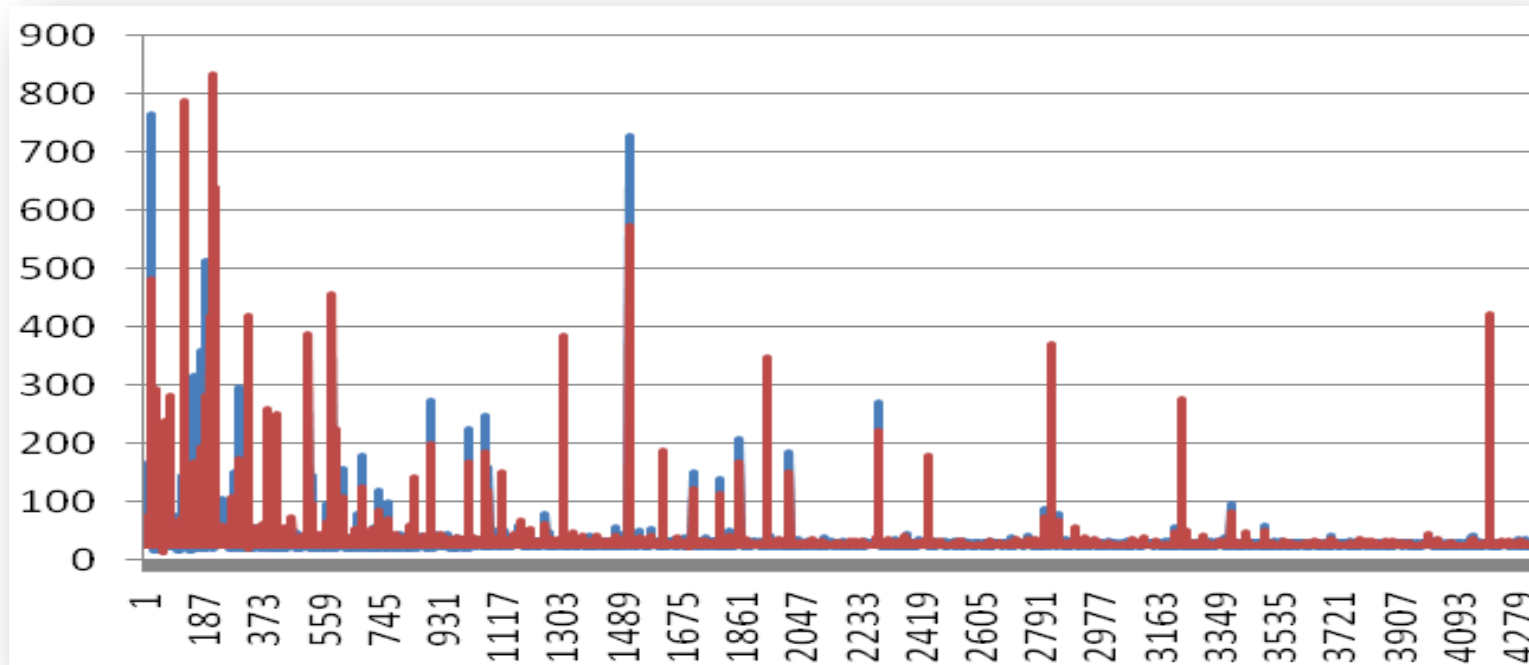


- Előnyös, ha
 - A problématerület gazdag történeti adatokban,
 - A megoldást meghatározó függvény ismeretlen vagy előállításuk költséges
 - Az célváltozót egymásra ható paraméterek írják le
 - Az adathalmaz bizonyos számú hibát tartalmaz

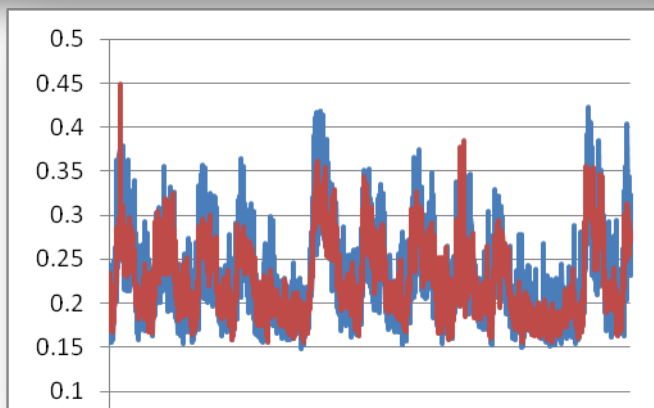
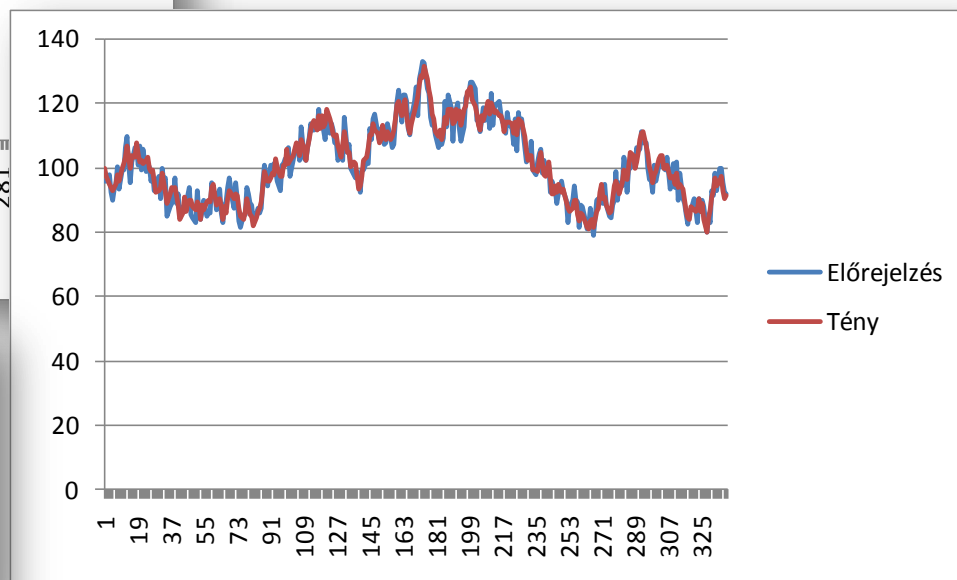
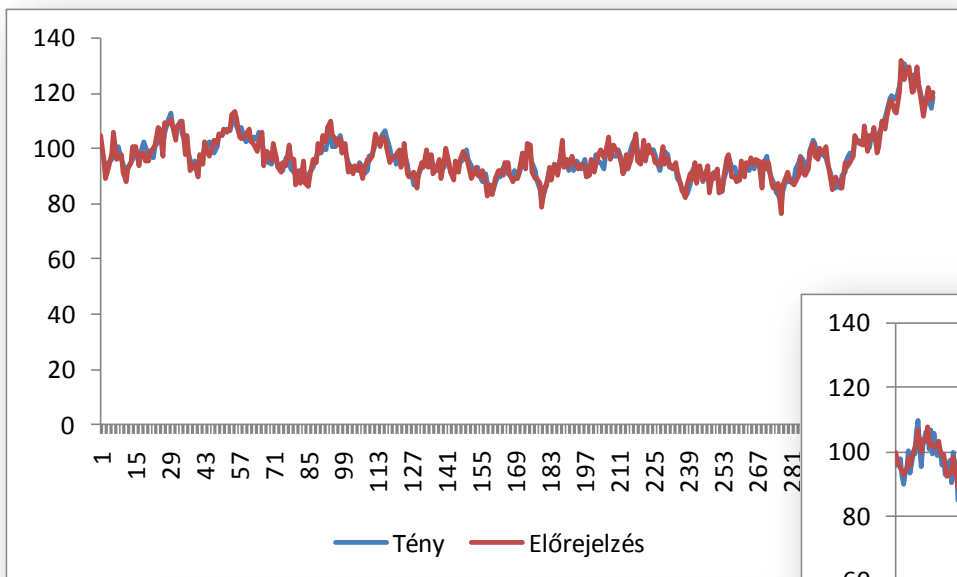


Jó előrejelző képesség

- 1 hónapnyi adat
- Illeszkedés: $r = 0,89$



További tesztek



ÉRTÉKELÉS



Tapasztalatok

- Tanító minta elemszáma
- Jól reprezentáló minták
- Feldolgozási kapacitás
- Modellek pontosítási igénye



Felhasználási lehetőségek

- SLM folyamat támogatása
 - Kampányok várható performancia hatásainak előrejelzése
 - Szűk keresztmetszetek meghatározása (időszakokra, vagy komponensre)
 - Szolgáltatás performancia képességek meghatározása
 - Kapacitás korlátok meghatározása
 - What-if próbák
- További adatkörök bevonása, modellépítés
- Alkalmazás méretezés
- Változások követése, visszamérés
- Kapacitástervezés



Üzenetek

Válság ellenére sem kell lemondani a szolgáltatásfejlesztésről

Innovációval kiemelkedő eredmények érhetőek el

Nem előfeltétel a drága eszköz

van olcsó **open-source** alternatíva
Pl. Rapidminer, Weka, stb.

Nagymennyiségű adat feldolgozása

Komplex elemzési lehetőségek

Újabb eszköz

Előrejelzések, trendelemzések biztosítása megalapozottabbá válik

A modell annyira pontos, mint az adat amivel használjuk

vegyük figyelembe az adattartalmat, és a lefedett adatköröket





 **BUDAPEST BANK**
A GE Money Bank tagja

 **CORVINNO**
Technology Transfer

 **KFKI**


**Köszönjük a
figyelmet!**



Várjuk a kérdéseket...³⁸