

Óbudai Egyetem

Neumann János Informatikai Kar



Prediktív analitika

Avagy üzleti jóslás tudományos alapon

Farkas László Zoltán





“A vonat és a ló közötti versenyhez hasonló történelmi pillanatban vagyunk, amikor az intuitív és tapasztalati (hüvelykujj) alapú szakértelmet meghaladják az adatokra épülő gépi modellek”

Bevezetés

Mióta ember él a Földön, kutatja a jövőt. Minket az érdekel, hogy mi várható a jövő marketing világában és a vezető vállalkozók köreiben. Az üzleti elemzés szerepe egyre inkább felértékelődik, ahogy a vállalatok mind nehezebben szereznek új vásárlókat, legyen szó akármelyik szektorról. Az elemzőeszközöket a vállalati vezetők leghatékonyabb fegyverei között tartják számon; különösen az előrejelzésre képes algoritmusokat becsülik sokra a bizonytalan gazdasági környezetben.

A Gartner piackutató cég becslése szerint a vállalatok 2015-ig 120 milliárd dollárt költenek hardverre, szoftverre és szolgáltatásokra csak azért, hogy javuljanak az elemzéseik, hogy jobban kihasználják a nagy adatban („Big Data”) rejlő lehetőségeket, hogy hatékonyabban menedzseljék pénzügyi műveleteiket, visszaszorítsák a csalásokat és megtudjanak valamit az ügyfélkapcsolatok következő fejlődési lépcsőjéről. Ehhez az úgynevezett prediktív analitikákat használják. – [3]

A prediktív analitika fogalma és működése

A prediktív analitika (továbbiakban PA) olyan statisztikai modelleket és egyéb empirikus módszereket jelent, amelyek empirikus előrejelzésekre szolgálnak, valamint a módszerek minőségi értékelésére, az



THE POWER TO PREDICT WHO WILL
CLICK, BUY, LIE, OR DIE

1. ábra „A jóslás hatalma” [12]

előrejelzések gyakorlati alkalmazhatóságára, azaz előrejelző képességének javítására törekszenek. „Eltekintve a gyakorlati hasznosság alapkritériumától, a PA fontos szerepet játszik a tudományos elméletben, a hipotézisek tesztelésében és relevanciavizsgálatában.” – [8]

Egyszerűbben megfogalmazva: a PA lényege, hogy minél több múltbeli adatból jelezzük előre (modellezzük) a jövőbeli várható viselkedést. Olyan matematikai, statisztikai, ökonometria módszerek összességét jelenti, amelyek ügyféladatbázisokon belül találnak összefüggéseket, és segítenek jobb üzleti döntések meghozatalában.

Minél nagyobb mennyiségű, ehhez szükséges adat áll rendelkezésünkre és minél pontosabban ismerjük az összefüggéseket az egyes vizsgálati elemek között, annál nagyobb megbízhatósággal határozhatjuk meg, hogy vizsgált szereplőink (játékosaink, vevőink, klaszteraink) miként reagálnak a jövőben, és közülük kik azok a kulcsszereplők, meghatározó elemek, akik megőrzése vagy befolyásolása különösen fontos.



2. ábra Prediktív analitika vázlatos folyamata [11]

A [8] anyag szerint Shmueli és Koppius 2010-es meghatározása szerint a PA két összetevőből áll:

- empirikus prediktív modellek (statisztikai modellek és egyéb módszerek, mint például adatbányászati algoritmusok), amelyek előrejelzést adnak a megfigyelések alapján,
- a módszerek értékelése, a modell prediktív ereje, azaz előrejelző képessége (prediktív pontossága). Az a modell képes generálni pontos előrejelzést, amelyben az „új” is értelmezhető időben, azaz a megfigyelések egy jövőbeni időszakában.

Nem minden analitika prediktív

Az analitika kifejezést nagyon széles körben és értelemben használják – azokra az elemzőeszközökre, amelyekkel nagy mennyiségű adatot lehet komplex matematikai és statisztikai algoritmusokkal vizsgálni. A PA egy olyan részterülete e módszereknek, amelyet jövőbeli trendek és viselkedésminták azonosítására használnak. – [8] [9]

Intuíció alapú döntések	Szabályozott döntéshozatal	Prediktív döntéshozatal
<ul style="list-style-type: none"> ❖ “Ösztön, előérzet, megérzés” ❖ “Hüvelykujj” szabályok ❖ Korábbi tapasztalaton alapuló döntések 	<ul style="list-style-type: none"> ❖ Tudás, szabályok és rutin az üzleti szabályokban ❖ A döntések hatékonyak és konzisztensek ❖ Objektív 	<ul style="list-style-type: none"> ❖ Pontos, optimalizált döntések tapasztalati mintákon ❖ Minden elérhető adat hasznosul ❖ Rugalmas, tény- alapú döntések ❖ Változó környezetben is robusztus- a legfrissebb adatok alapján újratanuló modellek reagálnak a trendekre

A Döntések típusai és legfőbb tulajdonságaik

A Prediktív elemzési megoldásokat a mindennapi működésbe beépítve a szervezetek kontrolálhatják a döntéseket, ezáltal üzleti céljaikat sikeresen teljesíthetik. Ezt úgy érik el, hogy múltbeli információkat felhasználva elemzik a jelenlegi helyzetet, majd előrejelzik a jövőt. – [10]

A múltbeli adatokat használják fel, hogy döntési folyamataikat optimalizálják, irányítsák és automatizálják. Az analitikák csendben megbújva teszik a dolgukat, elemzik az adatokat és küldenek szükség esetén riasztásokat, de nem minden monitoring vagy menedzsment eszközbe építik be őket. Sok eszköz segíti az adatok elemzését, de kevésben van olyan megoldás, amely megjósolná a jövőbeli teljesítményt a jelenlegi adatok alapján. – [12]



3. ábra A jóslás létrejöttének vázlatos folyamata [11]

Egy közkeletű hasonlattal élve a klasszikus üzleti intelligencia vagy BI¹ egy autó visszapillantó tükrében feltűnő képhez hasonló, míg a prediktív analitika vagy PA ahhoz hasonló, amit a szélvédőn át látunk. Más megközelítésben: a BI segít megmondani, hogy mi történt, miként működtetjük a céget, míg a PA azt, hogy mit kell majd tenni.

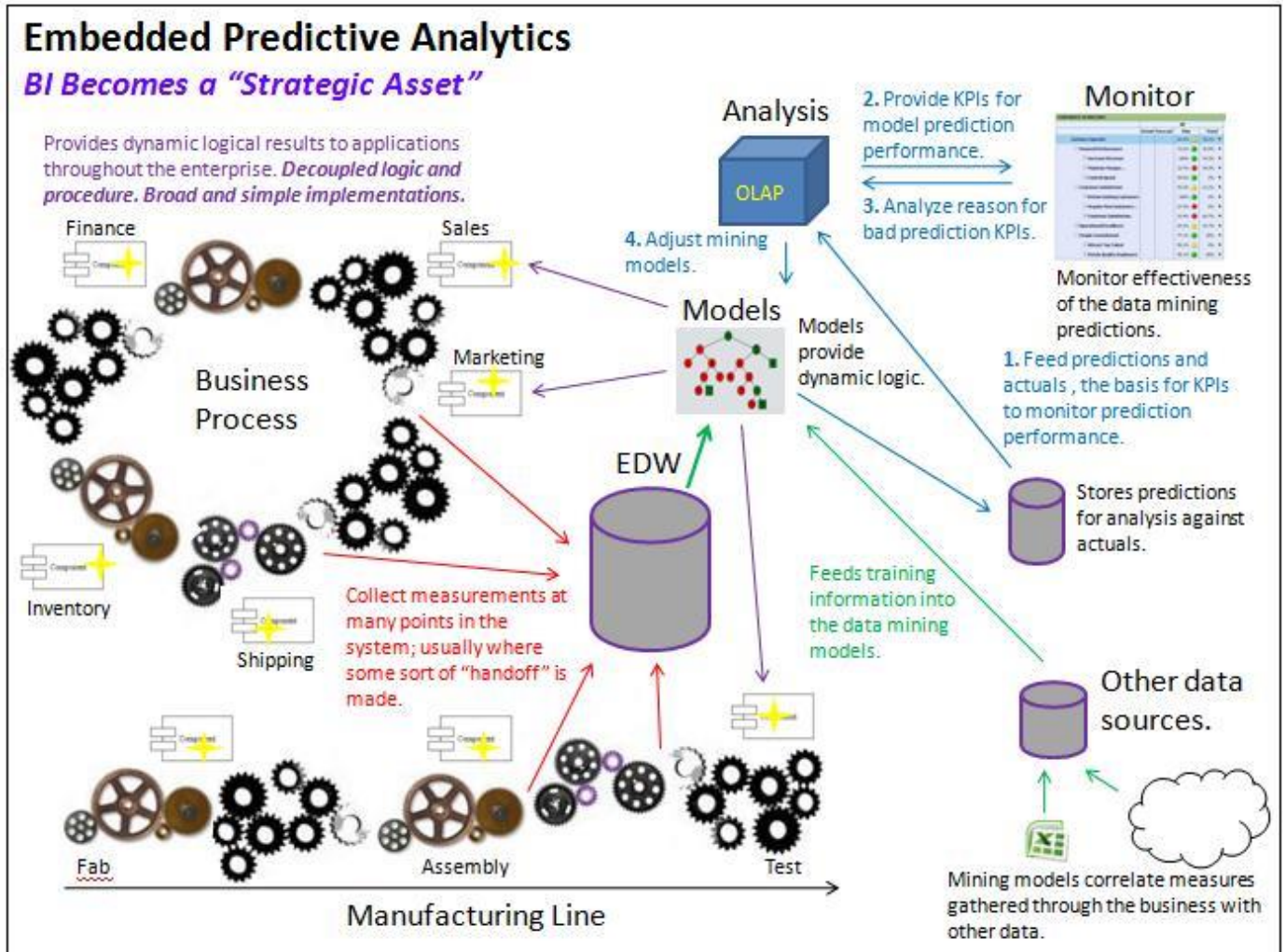
Az IT és a BI prediktív analitikái közötti különbség

A prediktív analízisnek két különféle alkalmazási iránya van: az egyik a BI, a másik az IT (Information Technology, Információs technológia, a gépi adatfeldolgozás eszközeinek és módszereinek összessége) számára.

Az üzleti intelligenciát szolgáló PA az adatbányászat és a matematikai elemzés eszközeivel dolgozik, hogy új, addig rejtett összefüggésekre leljen az adattömegben rejlő mintázatok között, és előre jelezze az üzleti trendeket például a fogyasztói aktivitásban, s az eredményei nem valósidejűek. [3]

Az IT-t támogató prediktív analitikák viszont valós időben készítenek hasonló elemzéseket. Ezt a technológiát az IT infrastruktúra minden szegletéből begyűjtött teljesítménymutatók elemzésére, a viselkedési minták azonosítására, és azoknak az IT adminisztráció felügyelői előtt rejtve maradó – anomáliáknak a meghatározására használják, amelyek befolyásolják a potenciális teljesítményt, illetve kapacitás problémákat okozhatnak. Az olyan dinamikus környezetben, mint a web-alapú alkalmazások az e-kereskedelemben, a virtuális környezetek vagy a számítási felhők, a prediktív analitikák nélkülözhetetlenek az IT számára. – [4]

¹ BI (Business Intel licence): olyan módszerek, fogalmak összessége, melyek a döntéshozás folyamatát javítják.



4. ábra Beágyazott Prediktív Analitikai rendszer felépítése és vázlatos működése [12]

Miben segíthetnek a prediktív elemzések?

A prediktív elemzésekből egy sor területen származhatnak üzleti előnyök. Sok adat és sok változó mellett a PA segít az adatok értékelésének automatizálásában, ami gyorsítja a döntéshozatalt. Az ügyfelek, beszállítók, és a többi partner, illetve a munkavállalók múltbeli viselkedéséből következtetni lehet arra is, hogy mit tesznek majd adott körülmények között a jövőben, ami segít meghatározni a legjobb lépéseket – minden egyes ügyfél és tranzakció esetében. – [11]

A döntési kockázatok gyakran attól függenek, hogy a döntés-előkészítés korlátozott erőforrásait mire használják. A PA az egységes döntési folyamatok lefektetését is segíti, s a döntések automatizálásával javulhat az ügyfelek kezelése. Ezt azonban számos előírás határozza meg, olykor már kezelhetetlenül sok. Ilyenkor a szabályhalmaz helyét a prediktív elemzési technikák vehetik át, amelyek egyszerűsíthetik magukat a szabályokat, de az általuk meghatározott folyamatokat is. [2]

Azok a döntések, amelyekre korábban órákat kellett szánni, akár másodpercek alatt is meghozhatók az előre kialakított banki termékektől az új telefon előfizetés ajánlatokon át a csalások valós idejű beazonosításáig. A „korai alkalmazók” már most ott tartanak, hogy a döntéshozás során akár a megelőző üzleti hét vagy nap teljes feldolgozott adatbázisára tudnak támaszkodni, ami a reakcióidő korábban elképzelhetetlen lerövidüléséhez, a piaci folyamatok és a pillanatnyi döntések majdnem valós idejű szintetizálásához vezet – ez különösen olyan időszakokban jelent nagy segítséget, amikor a szokásosnál gyorsabbak és gyakoribbak a nagy változások.

Ahol most vagyunk



5. ábra Grafikonok elemzése mindig könnyebb, mint a táblázatoké [4]

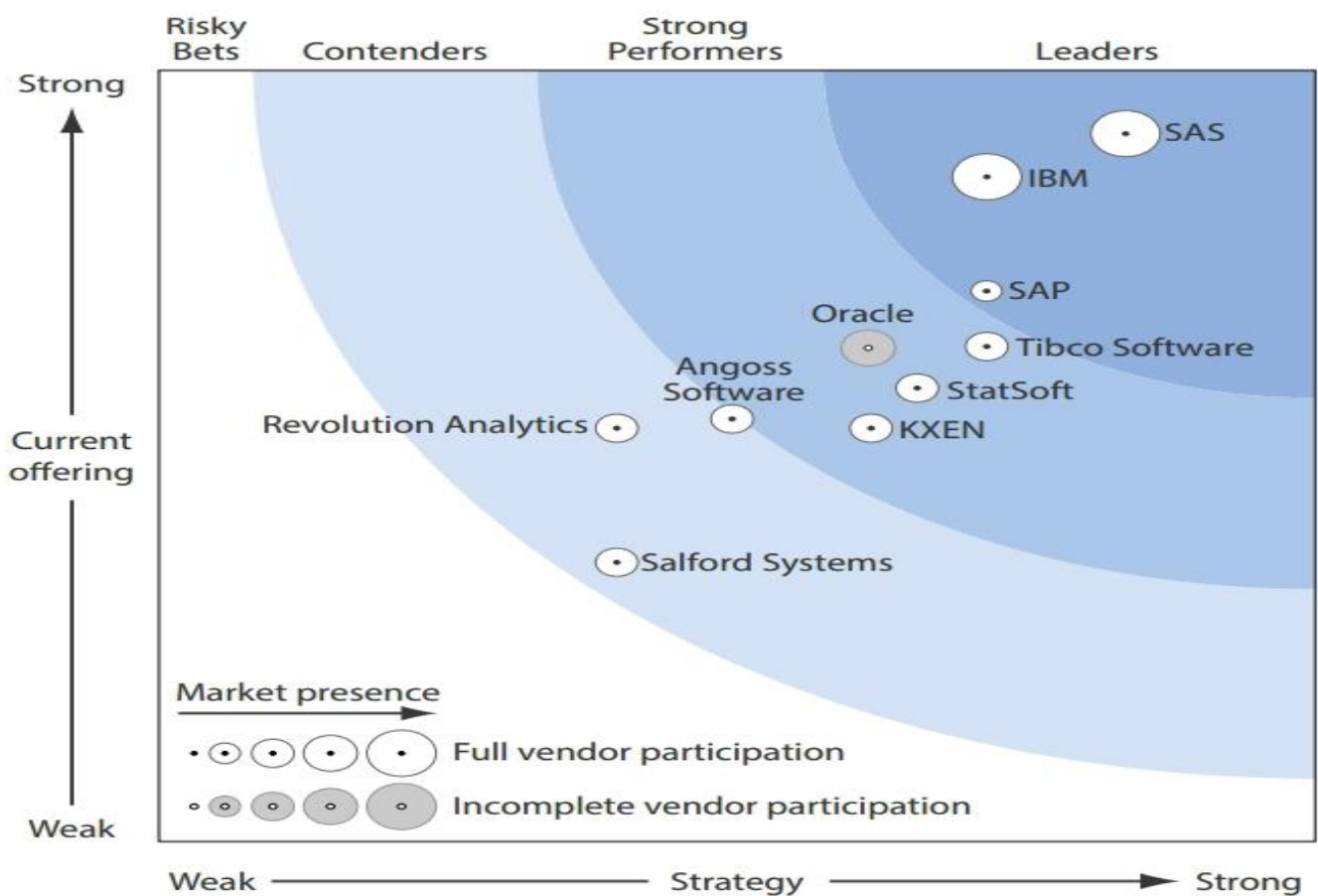
Az információ korunk gazdaságának mozgatórugója, legfőbb termelő ereje. A számítástechnikában adatnak (angol nyelven data) nevezzük a számokkal leírható dolgokat, melyek számítástechnikai eszközökkel rögzíthetők, feldolgozhatóak és megjeleníthetők. Ma

szinte felfoghatatlanul sok adat áll a rendelkezésünkre, amit egyre többen próbálnak meg rendszerezni és hasznosítani. A vállalkozások, a közösségi média és az egész internet exponenciálisan növeli az adatmennyiséget.

Az információkat reprezentáló adatok tárolására és elemzésére fejlesztett alkalmazásokba a legtöbb vállalat ma már jelentős tőkét fektet, hiszen a pontos és releváns elemzések növelik a vállalat működésének hatékonyságát, és támogatják a

stratégiaalkotás és megvalósítás folyamatát. - [3]

Az analitikai szoftverek iránti igény növekedése arra ösztönözte a fejlesztőket, hogy statisztikai eszközök és kimutatások minél szélesebb körét építsék az alkalmazásokba. Az innováció egyik legfőbb területe a prognózisok és előrejelzések készítésére alkalmas prediktív analitika - [2]. A prediktív analitikai világpiac jelenleg 2 milliárd dollár nagyságú, mely az IDC (ugyan csak egy piackutató cég) szerint a 3 milliárd dollár értéket fogja meghaladni 2017-re. Nem véletlen, hogy rengeteg nagy cég képviselteti magát valamelyik termékével ezen a piacon - [6].



6. ábra A Big Data és prediktív analitika vezető szállítói (2013. január 8) [4]

Már meglévő felhasználási területek (néhány példa)

Az üzleti világon kívül is rendkívül szélesek és változatosak a felhasználási területek, pl.

- **sport**
http://www.bi.hu/2011/09/09/prediktiv_analitika_teszi_szorakoztatobba_az_us_op_en_mecceit.html
- **politika**
<http://www.sajttaj.hu/index.php?page=tajekoztato&item=3046>
- **egészségügy**
http://www.sas.com/offices/europe/hungary/johir/klinikai_adatok.html
- **rendőrség**
http://hvg.hu/tudomany/20100727_prediktiv_analizes_szoftver_rendorseg_ibm
- **biztosítási csalások felderítése**
http://www.portfolio.hu/vallalatok/it/uj_megoldas_segitheti_a_biztositokat_a_csalasok_felderiteseben.184425.html

Merre tartunk?

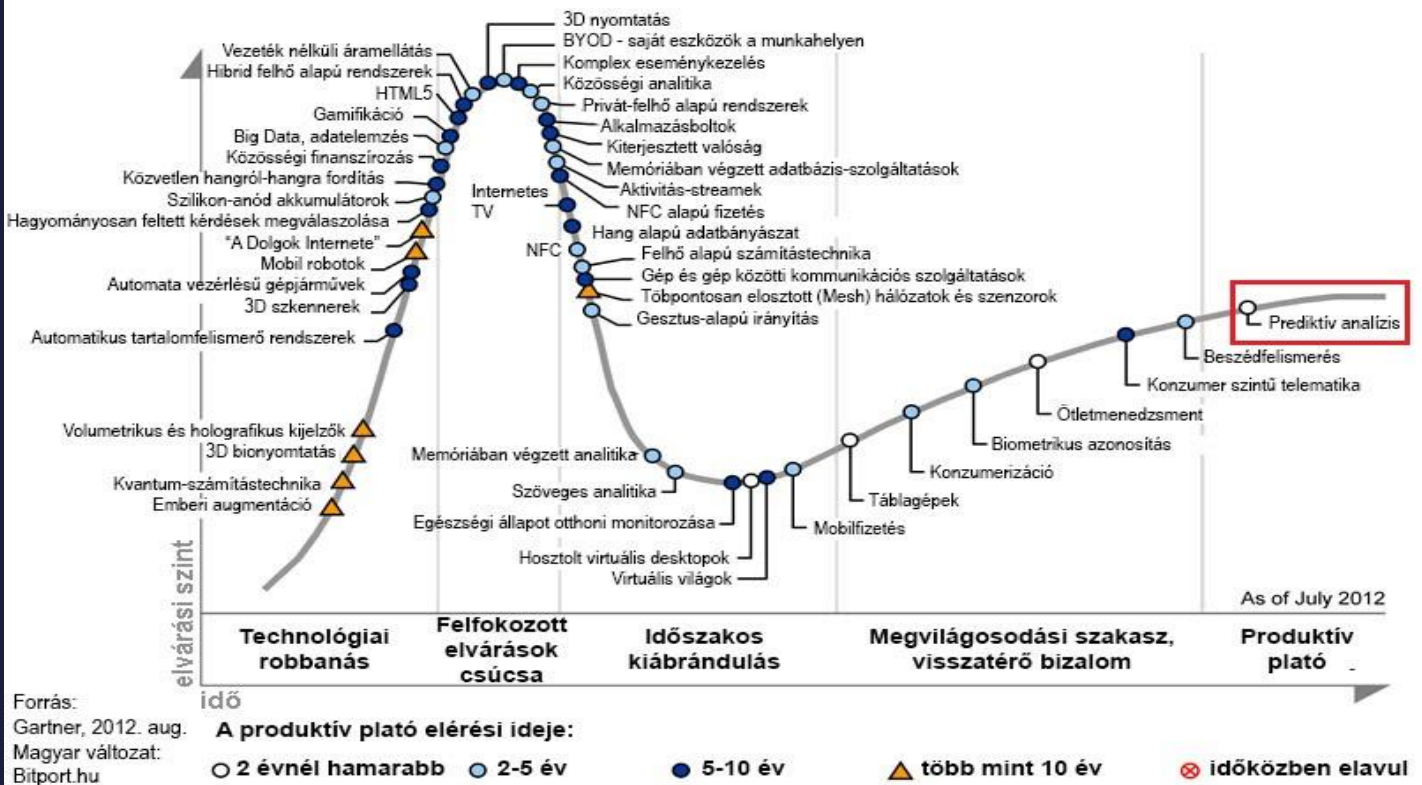
Az üzleti analitika korábban nagyrészt a strukturált adatok hozzáféréssel, kezelésével, tárolásával, modellezésével és optimalizálásával kapcsolatos eszközök, technológiák és megközelítések köré csoportosult- [11]. Ez változóban van, mivel a szervezetek egyre újabb és változatosabb adatforrásokba próbálnak betekintést nyerni. Ezeknek az eddig érintetlen adatforrásoknak a bevonása az elemzésekbe jelentős üzleti potenciállal járhat. Emellett jelentős a piaci hype a "big data" trend körül, ennek köszönhetően egyre több olyan megoldás lát napvilágot, amely a különböző adatforrásokból – így szövegből, e-mailekből, közösségi média bejegyzésekből, videóból vagy hangfelvételtől – is képes elemzéseket készíteni.

Előrejelzések

A Gartner három előrejelzést tett közzé, amelyet a BI csapatoknak érdemes figyelembe venniük, amennyiben a jövőt tervezik [3] [11] [12]:

- 2015-re az előre csomagolt analitikai alkalmazások mintegy kétharmadát fejlett analitikai funkciókkal, a Hadoop (<http://hadoop.apache.org/>) Big Data keretrendszerrel együtt fogják kínálni.

- 2015-ig világszerte 4.4 millió munkahelyet kell majd betölteni a Big Data elemzésében jártas szakemberekkel
- 2015-re az analitikai projektek 30 százaléka esetében az elemzések strukturált és nem-strukturált adatok kombinációjára épülnek majd.
- 2016-ra a vezető BI-szállítók 70 százaléka épít majd be megoldásaiba természetes nyelvi, illetve hangelemzést.



7. ábra Gartner Hype Cycle 2012 (<http://www.gartner.com/newsroom/id/2124315>)

Az 5. ábra negyedik oszlopába sorolt témakörök már kifejezetten jó helyzetben vannak: ezek túléltek az előző oszlop megpróbáltatásait, kiállták az idő és a piac próbáját, így újra felszálló ágba kerülve menetelnek a siker felé: a Gartner szerint ide sorolhatók a számítástechnikában végleges helyet találó táblagépek, az üzleti folyamatokban is használt konzumer eszközök gyakorlata, vagy épp a biometrikus azonosítási technológiák. Az ábrából kitűnik, hogy a közeljövő sztárja, a kutatás szerint, egyértelműen a prediktív analitika, amely az adatok elemzéséből kiindulva jövőbeli eseményeket, trendeket képes előrevetíteni.



SAP Predictive Analysis

(A következő leírás – beleértve a bemutatott ábrákat is, az sap.com és az scn.sap.com dokumentumai alapján készültek)

SAP BusinessObjects

Az SAP BusinessObjects Business Intelligence (BI) megoldások segítségével a felhasználók hatékony, megbízható adatokon és elemzéseken alapuló döntéseket hozhatnak. Az üzleti intelligencia megoldásokkal az összes felhasználó – az elemzőktől az alkalmi felhasználóig – hozzáférhet a számára szükséges információkhoz – és mindezt úgy, hogy minimálisan veszi igénybe az IT-t.

Az üzleti intelligencia megoldásokkal a felhasználók hozzáférhetnek a vállalat minden területéről származó információkhoz, navigálhatnak köztük, formázhatják, elemezhetik és megoszthatják azokat a különböző osztályok között.

- SAP BusinessObjects Explorer – „Ötvözi az intuitív információ-keresési és -feltérési funkcionalitást a nagy teljesítménnyel és skálázhatósággal”.
- SAP BusinessObjects Web Intelligence – „Webes alapú lekérdezés, jelentés-készítési és elemzési eszköz. Önkiszolgáló hozzáférést biztosít az üzleti információkhoz, lehetővé téve a felhasználók számára az adatok elemzését, a trendek beazonosítását, és segítségével az üzleti betekintések hatékony döntésekre válthatók.”
- SAP BusinessObjects Dashboard Builder – „Bármely üzleti felhasználó számára tervezhet interaktív keretet. A felhasználók rugalmasan testre szabhatják a keretet és dashboardokat, riportokat, egyéb tartalmakat hozhatnak létre, amelyek megfelelnek saját munkamódszereiknek. Az egyszerűsített vizuális megjelenítéshez való könnyű hozzáférés révén a felhasználók számára rendelkezésre állnak a szükséges információk és ezért a nagyobb értéket teremtő

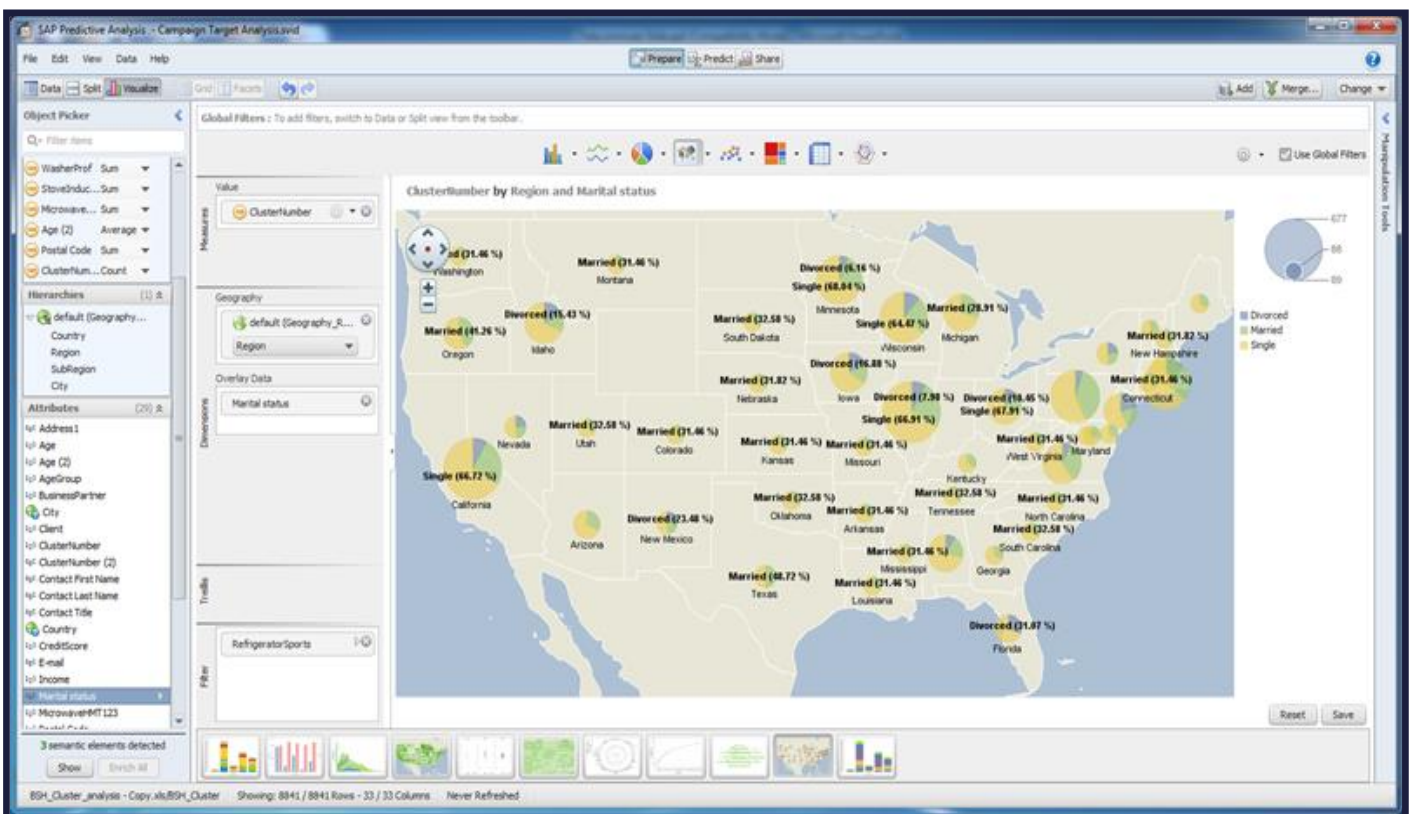
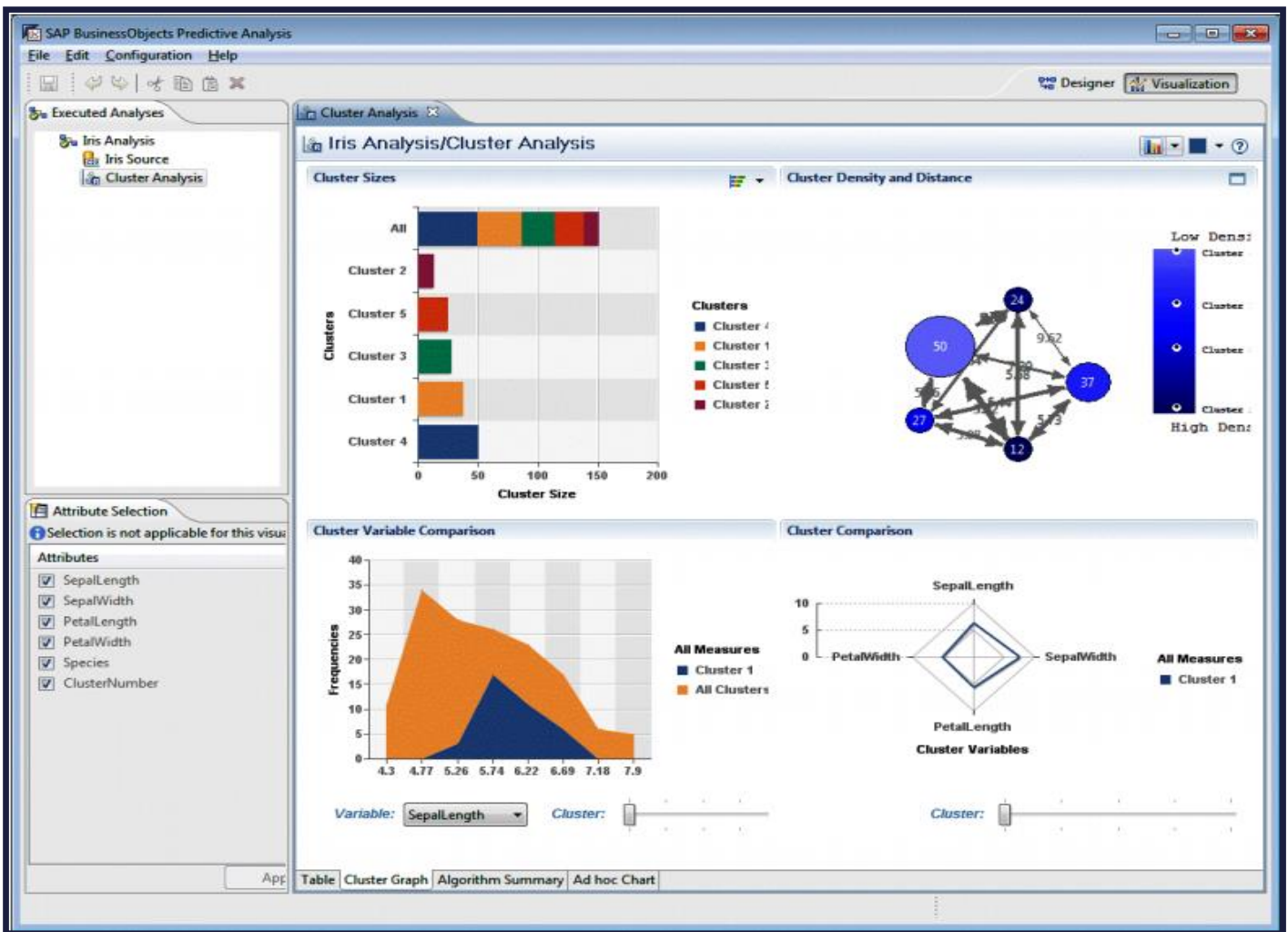
tevékenységekre koncentrálhatnak.”

- Crystal Reports Software – „Hatékony, dinamikus és végrehajtható lépéseket leíró jelentéskészítési megoldás, amellyel megtervezheti, kiaknázhathatja, vizualizálhatja és elkészítheti riportjait akár az interneten keresztül, akár a vállalati hálózatban.”
- SAP BusinessObjects Live Office – „Integrálja az üzleti intelligenciát. Megbízható, frissíthető adatok ágyazhatóak be a dokumentumokba, táblázatokba, prezentációkba, amelyeket megoszthat kollégáival is.
- SAP BusinessObjects Voyager – „A szoftver segítségével az elemzők dolgozhatnak az OLAP adatokkal, megválaszolhatnak üzleti kérdéseket, trendeket fedezhetnek fel és elemzéseiket megoszthatják másokkal. Az elemzők többdimenziós adatsomagokat nézhetnek meg és betekintést nyerhetnek a szervezeten belül megtalálható OLAP adatok teljes skálájához való hozzáférés és megtekintés révén.”

Néhány példa a program funkcióiról és képességeiről:

The screenshot displays the Clementine 12.0 interface. On the left, a workflow diagram shows data from 'BASKETS1n' being processed through several '11 campos' and '11 fields' nodes. In the center, a scatter plot titled 'Malla de 11 Campos : Absolutos' shows a distribution of data points for various categories like 'dairy', 'cannedveg', 'fish', 'beer', 'freshmeat', 'frozenmeat', 'wine', 'softdrink', 'confectionery', 'cannedmeat', 'dairy', 'fish', and 'fruitveg'. On the right, a table window titled '11 campos' displays association rules with columns for 'Consecuente', 'Antecedente', '% de soporte', and '% de confianza'. Below the table are tabs for 'Modelo', 'Resumen', and 'Anotaciones'. The bottom of the interface features a toolbar with various modeling and analysis tools.

Consecuente	Antecedente	% de soporte	% de confianza
cannedveg	freshmeat	3,0	96,67
beer	frozenmeat	3,1	93,55
frozenmeat	cannedveg	4,0	90,0
cannedveg	cannedmeat	4,5	88,89
beer	fruitveg	3,3	87,88
beer	cannedveg	16,7	87,43
beer	freshmeat	4,6	86,96
frozenmeat	cannedveg	2,3	86,96
frozenmeat	dairy	2,3	86,96
frozenmeat	frozenmeat	2,3	86,96
cannedveg	beer	17,0	85,88



Felhasznált irodalom és források:

- [1] BI Portál-Az üzleti intelligenciáról cikkei alapján
<http://www.bi.hu>
- [2] FICO összefoglaló cikkei
<http://www.fico.com/en/Communities/Pages/Predictive-Analytics.aspx>
- [3] HVG online cikkei alapján
<http://hvg.hu>
- [4] SAS Hungary weblapja
<http://www.sas.com/offices/europe/hungary/>
<http://blogs.sas.com>
- [5] SAP dokumentumai és reklámanyagai
<http://www.sap.com/index.epx>
<http://scn.sap.com/welcome>
<http://www.saphana.com/community/learn/solutions/predictive-analysis>
- [6] Computerworld cikkei
<http://computerworld.hu>
- [7] Üzleti Döntéstámogató Prediktív analitikai alkalmazások
<http://www.spss.hu/>
- [8] Tessényi J. – Kazár K. (2012): Szerencsejáték-vásárlási szokások vizsgálata „prediktív analitika” segítségével. Statisztikai Szemle, 90, 7-8, 677-695. o.
http://www.ksh.hu/statszemle_archive/2012/2012_07-08/2012_07-08_677.pdf
- [9] MarketingGPS.hu
<http://blog.marketinggps.hu/cikkek/-latni-es-megvaltoztatni-a-jovot-a-21-szazad-legszexibb-foglalkozasa/99>
- [10] Wikipédia
http://en.wikipedia.org/wiki/Predictive_analytics
- [11] Predictiveanalyticsworld (témával foglalkozó szakportál)
http://www.predictiveanalyticsworld.com/predictive_analytics.php
- [12] Northwestern University School of Continuing Studies (egyetem nyilvános anyagai)
<http://www.scs.northwestern.edu/program-areas/graduate/predictive-analytics/>